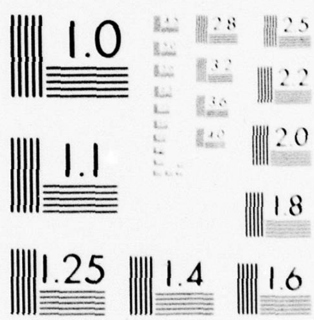
AD-A057 650 ILLINOIS UNIV AT URBANA-CHAMPAIGN COORDINATED SCIENCE LAB F/G 14/5
ON COMPUTER STEREO VISION WITH WIRE FRAME MODELS.(U)
DEC 77 D J BURR DAAB07-72-C-0259
UNCLASSIFIED R-805 NL

1 OF 2
AD
A057 650



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

REPORT R-805 DECEMBER, 1977 UILU-ENG 77-2251

CSL COORDINATED SCIENCE LABORATORY

ON COMPUTER STEREO VISION WITH WIRE FRAME MODELS

DAVID JOSEPH BURR

APPROVED FOR PUBLIC RELEASE. DISTRIBUTION UNLIMITED.

UNIVERSITY OF ILLINOIS - URBANA, ILLINOIS

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) ON COMPUTER STEREO VISION WITH WIRE FRAME MODELS	5. TYPE OF REPORT & PERIOD COVERED Technical Report
	6. PERFORMING ORG. REPORT NUMBER R-805; UILU-ENG 77-2251
7. AUTHOR(s) David Joseph Burr	8. CONTRACT OR GRANT NUMBER(s) DAAB-07-72-C-0259
9. PERFORMING ORGANIZATION NAME AND ADDRESS Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Joint Services Electronics Program	12. REPORT DATE December, 1977
	13. NUMBER OF PAGES 117
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report) UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Computer Stereo Vision
3-D Scenes
Scene Reconstruction

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report treats two major subproblems in computer stereo vision -- ~~(1)~~ that of reconstructing 3-D scenes from stereo sets of images and ~~(2)~~ e automatic recognition of 3-D scenes.

Two techniques are presented for scene reconstruction. The first, or multiple view method, utilizes the combined information from bulk correlation and three or more stereo images to construct three-dimensional edge features or structures. The structure is obtained by projecting into space a piecewise-linear representation of intensity edges obtained from one → next page

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT (continued)

of the images. The second technique utilizes a narrow angle pair (2-3 degrees) of images and symbolic correlation to ensure matching reliability and efficiency in the construction of edge depth maps of scenes. A new technique of dynamic smoothing of edge contours is presented which permits accurate triangulation at narrow viewing angles, while preserving the integrity of sharp corners. Also, two new techniques are presented for piecewise approximation of 3-D and 2-D digital contours with circular arcs. In both stereo techniques objects with prominent edges are preferred, but no other restrictions are made on surface shape. In this sense the work represents a major advancement over previous techniques for stereo vision, which generally restrict objects to have plane faces or simple curved surfaces.

In the second category, or recognition, a technique is described for matching a 3-D scene reconstruction containing piecewise-linear edges (e.g. constructed by the multiple view method) to a stored wire frame model, based on utilization of 3-D features and geometric constraints. Techniques are suggested for implementing efficient search in occluded and cluttered scenes, and in cases where there are many models.

LEVEL II

12

14 R-815, UILU-ENG 77-2251

9 Doctoral thesis,

6 ON COMPUTER STEREO VISION
WITH WIRE FRAME MODELS.

by

10 David Joseph Burr

11 Dec 77 12 127 p.

15 DAAB07-72-C-0259

This work was supported in part by the Joint Services Electronics Program (U.S. Army, U.S. Navy and U.S. Air Force) under Contract DAAB-07-72-C-0259.

78 08 15 028

097 700

ON COMPUTER STEREO VISION WITH WIRE FRAME MODELS

BY

DAVID JOSEPH BURR

B.A., Franklin and Marshall College, 1968
M.S., University of Illinois, 1971

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 1978

Thesis Advisor: Professor Robert T. Chien

Urbana, Illinois

ON COMPUTER STEREO VISION WITH WIRE FRAME MODELS

David Joseph Burr, Ph.D.
Coordinated Science Laboratory and
Department of Electrical Engineering
University of Illinois at Urbana-Champaign, 1978

This thesis treats two major subproblems in computer stereo vision -- (1) that of reconstructing 3-D scenes from stereo sets of images and (2) automatic recognition of 3-D scenes.

Two techniques are presented for scene reconstruction. The first, or multiple view method, utilizes the combined information from bulk correlation and three or more stereo images to construct three-dimensional edge features or structures. The structure is obtained by projecting into space a piecewise-linear representation of intensity edges obtained from one of the images. The second technique utilizes a narrow angle pair (2-3 degrees) of images and symbolic correlation to ensure matching reliability and efficiency in the construction of edge depth maps of scenes. A new technique for dynamic smoothing of edge contours is presented which permits accurate triangulation at narrow viewing angles, while preserving the integrity of sharp corners. Also, two new techniques are presented for piecewise approximation of 3-D and 2-D digital contours with circular arcs. In both stereo techniques objects with prominent edges are preferred, but no other restrictions are made on surface shape. In this sense the work represents a

major advancement over previous techniques for stereo vision, which generally restrict objects to have plane faces or simple curved surfaces.

In the second category, or recognition, a technique is described for matching a 3-D scene reconstruction containing piecewise-linear edges (e.g. constructed by the multiple view method) to a stored wire frame model, based on utilization of 3-D features and geometric constraints. Techniques are suggested for implementing efficient search in occluded and cluttered scenes, and in cases where there are many models.

DEDICATION

The author wishes to dedicate this work to his loving parents, Jennie Catherine Burr and Joseph Henry Burr, both of whom died unexpectedly during the course of the research. They were a constant source of encouragement to him.

ACKNOWLEDGEMENT

The author wishes to thank his advisor, Professor Robert T. Chien, for his support and general guidance.

The author would like to thank Professors David L. Waltz and Richard M. Brown for serving on the thesis committee and for helpful advice. Thanks are extended to Professor Wesley E. Snyder for camera hardware and to Dr. Vincent C. Jones for his picture handling routines. The author also wishes to thank his colleagues at the Coordinated Science Laboratory for helpful discussions and, in particular, Charles J. Jacobus for programming advice and camera maintenance.

In the "last but not least" category are the author's loving wife Barbara and daughter Lori, who showed much patience and understanding through the years. The author is grateful to Professor Leonard V. Cherry of Franklin and Marshall College, who introduced him to the exciting world of scientific research.

TABLE OF CONTENTS

1. INTRODUCTION

Few persons would deny the usefulness of a computer system which could accept the visual output of a television camera and interpret the image. Its impact in industry alone would be overwhelming, since the visual inspection task is a common problem, requiring tedious attention by humans. Automated manufacturing systems employing computer controlled manipulators typically work blindly and have little or no ability to recover from errors. Visual feedback would allow error recovery, visual servoing, and unstructured input to the assembly process. Applications abound in hostile environments, extraterrestrial exploration, and earth resources technology. Automation of routine tasks such as counting, sorting, and recognition of cells and fine particles would prevail in microscopy, medicine, and industrial quality control. Security control and intrusion monitoring would also benefit. Implications of scene analysis as applied to data compression and bandwidth reduction problems are overwhelming.

Acknowledging the difficulty of 3-D scene interpretation, most early research focused on line drawings of simple objects. Currently there is strong emphasis on real images of complex-shaped objects in cluttered scenes. In conjunction with this trend, new and more powerful techniques are being developed, including the use of

multisensory information such as range, color, and texture. Many feel that range information is a prerequisite to successful understanding of complex three-dimensional shapes by machines.

Most of the early work dealt with interpretation of 3-D scenes from a single image. Since a great deal of structural information is lost in a projected image, the scenes had to be highly restricted in domain, even to the extent of excluding curved objects. Therefore the early work emphasized projective features of simple objects, and resulted in much information about vertex types and relationships in visual scenes. In other work 3-D models were employed, but comparisons were made to model projections. The point is that much of the early work dealt essentially in two-dimensional ideas.

3-D objects, however, are best characterized by 3-D prototypes, and matching of such structures is best accomplished in the spatial rather than the projection domain. If nothing more, search efficiency is increased due to absolute knowledge of lengths, angles, and positions. Furthermore, domain limitations imposed by working in projected images no longer need apply, and complex unrestricted shapes should be treatable.

Thus it is felt that significant advances in computer vision will necessarily result from research in three-dimensional feature construction. This will allow

unrestricted shapes to be modeled and recognized by machines working on real images. This work contributes several new techniques in this area.

The purpose of this work is to investigate 3-D feature extraction from images of objects with arbitrary curved edges, and to study the comparison of 3-D features for model-based object recognition. The domain consists of common objects of rigid form and no significant surface textures. The overall goal is the understanding of stereo image perception in relation to the design of automated scene analysis systems. Thus the emphasis of this work is on image understanding as opposed to image processing.

The strength of this work in contrast to much prior work in computer vision is its full consideration of the problems associated with real images, the compatible treatment of both low and high level aspects of the vision problem, and the ability to deal with complex curved shapes.

2. SUMMARY OF CONTRIBUTIONS

The main contributions presented here are -- (1) A multiple view method for enhancing bulk correlation peaks, thus permitting reliable matching of low information image areas; (2) A narrow angle method using edge smoothing and spatial edge continuity for computation of edge depth maps; (3) A new nonlinear or dynamic technique for smoothing digital contours while minimizing corner rounding; (4) Several new techniques for iterative fitting of circular arcs to two- and three-dimensional contours; and (5) A technique for comparing three-dimensional geometric structures, for use in model-based recognition of objects in 3-D visual scenes. The contributions are generally in the area of new techniques for constructing and comparing 3-D features from visual scenes for use in automated scene understanding systems.

In the work on stereo image comparison for extraction of structural features, it was felt important to consider the requirements of shape representation and feature selection as well. As a result two different approaches have resulted, each based on different such criteria. The first, or multiple view method (Chapter 5), requires three or more simultaneous views, and exploits redundancy in the set of views to enhance the matching capability of conventional bulk correlation at low information windows.

Though usable alone it is described in conjunction with a technique for building a 3-D scene description from a 2-D segmented image. The efficiency is gained through use of heuristics to limit search normally required with a cross-correlation operator.

The second method (Chapter 7) is developed around the assumption that it may be desirable to represent 3-D scenes with symbols less wordy than linear segments. The requirement to limit combinatorics in matching to model features suggests piecewise circular descriptions or other quadratic primitives for representing 3-D edges. The desire to first obtain 3-D depth maps of edges puts more severe requirements on the efficiency of the image matching process, since many depth values must be computed. The concepts of narrow angle stereo pairs, symbolic feature matching, and 2-D and 3-D continuity are indeed powerful and serve as the basis for this approach. Because of a conflicting requirement for accuracy of triangulation, the narrow angle approach must be augmented by an additional technique to reduce edge noise and quantization. This is embodied in a new dynamic smoothing technique which greatly reduces noise while minimizing the deterioration of corners.

Some contributions are also made in the area of contour approximation with circular arcs (Chapter 8). Two methods are presented which are heuristic in nature, and in one case utilize the fact that edge points are connected and are

roughly uniformly spaced.

Determination of a 3-D symbolic map based on piecewise circular or linear primitives can be considered as an intermediate step in the computer vision problem. One would like to develop techniques to extract meaning from the relationships existing among these symbols in particular scenes. Object-independent approaches to this problem have resulted and they center upon vertex based ideas (Guzman (1968), Waltz (1972)). Our approach is that of iconic modeling of particular objects and associated techniques for matching such models to the 3-D scene. The notion is motivated by the fact that there are often requirements in practice for systems that can identify scenes consisting of a limited set of objects (i.e. industrial assembly, inspection, etc.). In addition, because such symbolic maps are necessarily incomplete and locally erroneous, means are needed for disambiguating them on the basis of what is known or expected about possible objects. The contribution presented here is a technique for matching incomplete 3-D line constellations with wire-frame object models (Chapter 6). This is done by exploiting geometric constraints between 3-D edge features in the scene and in the a priori encoded models. The program assigns a figure of merit for the existence of a particular object in a part of the scene. Strategies for selecting a plausible scene interpretation based on these figures of merit are discussed.

3. RELATED WORK

Because of subproblems present in stereo vision as seen here, the related work has been divided into several categories. They are discussed separately under the various subheadings of Chapter 3. The reader who is acquainted with the history of work in stereo and monocular computer vision may find it desirable to proceed directly to Chapter 4.

3.1 General.

Although there has been considerable previous work concerning stereo correlation of image pairs and much work on extraction of shape features from monocular images, little has yet been accomplished on the composite problem of 3-D feature determination. Because of possible tradeoffs between them, the problems should be studied jointly. In addition, there is genuine need for research on real rather than contrived images, since problems associated with real images have in the past not succumbed well to extensions of work on perfect drawings.

Modeling of rigid shapes with geometric constructs has received some attention in the light of computer vision. One can classify past work into categories based on 2-D and 3-D approaches for modeling, and for image feature extraction. Grape (1973) used 2-D approaches to both problems while dealing with simple polyhedral shapes. 3-D

modeling with 2-D feature extraction was treated by Roberts (1965), and is generally accepted as the first serious work in computer vision. However, the combination 3-D/3-D has not been studied in the general sense except for some work by Falk (1970). His approach to depth extraction was based on the restrictive assumption of known intersecting planes. In addition, strong assumptions about the presence of vertical edges prohibit extension beyond the polyhedral domain. Some recent work by Baker (1975) deals with the problem of constructing a surface description of a solid curved object, or learning by looking. The work emphasizes model building as opposed to recognition in scenes.

Hill climbing has been attempted in various forms to solve the problem of model-scene correspondence (Hemami et al. (1975), Barrow et al. (1977)). Unfortunately, such approaches do not circumvent the problem of determining correspondences between the model and the scene, and heretofore only simple techniques have been tried. The fundamental problems of feature selection and matching remain open. In addition, one has the added problems due to local extrema, requiring extensive search and separate means for deciding when particular extrema are significant. Simple hill climbing is adequate only when an initial correspondence is sufficiently close to the correct one, and thus may be useful only for fine tuning of proposals made by more sophisticated means. Furthermore, no work exists to my knowledge on hill climbing to match a 3-D structure with

another 3-D structure. This might be attractive since the task of shape feature matching could be simpler, relating features in the same dimension.

Problems related to local extrema perhaps could be eliminated by appropriately blurring the error function at different stages of iteration. This might be aided by incorporating higher level features as in our model matching scheme, and by including information regarding the discriminability of particular features for indicating rotational and translational shifts. Furthermore, evidence exists for the need to fine tune (de-blur) error functions with time, since the process hopefully converges, and correctly so if the dynamics are treated properly. The ideas are not unlike relaxation labeling techniques, which are discussed in Section 6.4.1.

The Fourier descriptor approach is another essentially 2-D/2-D approach to matching boundary shapes of a projected object. It has also been used in character recognition studies. The problems of this approach are the large number of views that must be modeled for each object, and the difficulty of treating partial shape descriptions, which is essential for occluded scenes.

Past work in stereo image comparison can be classified also into wide and narrow angle approaches. Because of conflicting requirements between subproblems in stereo correlation, certain tradeoffs must be made. The conflicts

consist of determination of matched pairs and triangulation accuracy. If the two images differ only slightly, say by slight change in view aspect, then matching is simplified. However, triangulation suffers due to near parallel intersection of rays coupled with image noise. The triangulation problem is normally solved by introducing larger disparity between views, however at the expense of requiring greater sophistication in feature matching. Another basic limit to large viewing angle is the decreasing visual overlap between the two scenes.

The choice to use global versus local information in the comparison of features is determined to a certain degree by the technique used for comparisons. When bulk techniques are used window size is limited due to distortions arising from viewing aspect and perspective projection, and is thus local. However, feature extraction followed by symbolic matching is not so restrictive, and global information is more easily incorporated in the matching. This is a strong argument in favor of symbolic matching aside from its inherent speed advantage. However, in real images connectivity of global shape features is not easily exploited due to missing and extraneous segments.

It is held here that the approach to computer vision involving 3-D features has great promise in being successful on real object scenes. Certainly the more information that can be brought to bear in a knowledge-based system, the

better the chances of success. But even more, the fact that real objects are essentially three-dimensional allows the modeling of highly relevant shape features, and thus straightforward techniques for comparing them. In addition, the use of powerful 3-D geometric constraints would allow filling in of missing information and elimination of spurious edges, based on geometry alone. Thus segment connectivity need not be enforced in model matching of rigid objects.

3.2 Stereo Image Comparison and Depth Ranging.

3.2.1 Bulk Correlation.

Although related work in this area emphasizes the construction of depth maps of textured scenes, such techniques are also useful for building higher-level 3-D features. Such works are generally highly successful, and not surprisingly so, since textured areas exhibit strong locally discriminating patterns for use in matching. A thorough treatise on the subject of bulk correlation as applied to matching of stereo image pairs is that of Hannah (1974). Ideas are discussed in relation to window specification, search reduction with and without camera models, continuity implementation, and detection of unmatchable features. Two algorithms for implementing many of these ideas are described. The work serves as a good reference for equations involved in correlation and

projective geometry.

Applications using bulk techniques include Levine et al. (1973) and O'Handley (1973), in which complete depth maps are computed for a simulated Martian terrain. A device for computing cross correlations is described. In spite of the savings in time, 30 minutes or more is required for computing depth maps. One would learn from this that bulk correlation methods should be used carefully, perhaps guided by higher level programs to isolate matching to essential places.

Nevatia (1976) solves the narrow angle triangulation problem by tracking features while a scene is rotated. The matching is easy since it is done between incrementally shifted pictures. The narrow angle method described here compares well with his accuracy, requiring only two pictures instead of many. Other applications-oriented works include Pingle and Thomas (1975) in which Nevatia's work is extended by comparing corner features. Quam (1971) and Quam and Hannah (1976) treat problems associated with satellite imagery and geometric distortion. Other treatments of geometric distortion include Markarian et al. (1973) and Wong et al. (1973).

Two iterative techniques for matching images include a relaxation labeling or cooperative method (Marr and Poggio (1976)), and a hill climbing method (Mori et al. (1973)). Marr and Poggio match random dot stereograms by a nonlinear

relaxation process. The semantic information of uniqueness and continuity are used exclusively with a very local one pixel correlation measure. It treats the problem of overcommitment to a particular interpretation by allowing all possible interpretations to grow in parallel. During this growth the various interpretations (labelings) influence each other in positive and negative manner, and their outcome is modified by a nonlinear decision function. The finite state model they describe converges to final states which are appropriate depth maps for random dot stereograms. Another iterative technique is the prediction-correction method of Mori et al. (1973). Though resembling hill climbing more than relaxation labeling, it attempts to predict terrain disparities using conventional bulk correlation methods. The predicted disparities are locally smoothed, and the differences between a disparity-shifted image and the original are determined. Differences define a correction to be applied to the estimated disparities. This continues until the differences (error) are within prescribed limits.

3.2.2 Symbolic Correlation.

The author defines symbolic correlation as any matching process which compares properties of derived features (e.g. edges, lines, curves, or vertices), as opposed to raw intensity patterns (bulk correlation). Symbolic correlation is often implemented as a linear or nonlinear threshold

function having the respective feature differences as arguments (e.g. edge orientation, vertex type, arc curvature, etc.). When the feature is naturally discrete such as vertex type, then it is desirable to impose some scalar order on the symbols in the set (see Ganapathy's (1975) approach to this problem).

The symbolic approach is preferred when wide angles are used, since simple cross correlation fails. Works using this approach relate more closely to this thesis in the sense that high level features are treated. However, since extended rather than local features are often used, too great an emphasis is generally put on the need for similar segmentations in image pairs. Thus object domains are usually restricted to polyhedra or simple curved surfaces. This thesis does not make such restrictions.

Work dealing with plane-faced solids is first presented. Perkins (1970) assumes a pair of line drawings of polyhedra as input. Relying strongly on the notion of vertex connectivity, he matches views by searching a tree of all possible matches, pruning when inconsistencies develop. His notion of the correct match is that one which forms the maximally connected map of all possible maps. Both image-derived and hand-drawn scenes are tested. Ganapathy (1975), on the other hand, acknowledges that matching of wide angle views may draw upon diverse strategies to resolve ambiguities. He describes seven matching heuristics and

attempts to order them according to local and global usefulness. Examples are shown for simulated line drawings of up to ten polyhedral objects, and for real images of up to three polyhedra.

Works using symbolic techniques in the curved object domain consist of the following: Baker (1975) employs a set of sequential views to build surface descriptions of smooth objects with arbitrarily curved surfaces. He matches view pairs of a single object by comparing arc features at curvature irregularities of region boundaries. Real images are used in the work. In some recent work Shapira (1977) uses a vertex ordering technique aided by three-view redundancy to match real images of objects with simple quadratic surfaces. The three views are taken at very wide angles about the scene. Based on the global nature of vertex ordering, some missing scene edges are proposed.

Underwood and Coates (1975) deal with the matching problem but not as relates to stereo comparison. They use a projective invariant, the "cross ratio" (Duda and Hart (1972)), to match polygonal faces of polyhedral images. In this way they build a surface description graph of an object from several single views. Though the cross ratio can be generalized to aid the matching of curved edges, the problem of determining corresponding points between the curves appears to be nontrivial.

In summary, little has been done quantitatively on comparing real images of complex curved objects to extract shape features, without making highly restrictive assumptions about surface shape.

3.2.3 Other Methods.

Another work which attempts to build 3-D descriptions of objects is that of Baumgart (1974). He computes volumetric structures by intersecting projection cones of an object obtained from a sequence of views. Ambiguity normally associated with stereo methods is absent, since left-right ordering is enforced when silhouettes are intersected. Errors thus get introduced, such as the filling in of certain concave portions of objects. Baumgart also treats geometric modeling in a general sense and describes a "winged edge" data structure for wire frame models. The emphasis of his work is graphic modeling.

Other works (Della Vigna and Luccio (1970), Shapira (1974)) exist in wide angle stereo, but their purpose is mainly to formalize some problems, and they do not contribute to resolving the match ambiguity.

A number of methods for direct extraction of depth exist, using time of flight of a light pulse (Duda and Nitzan (1976)), light beam triangulation (Agin (1972), Fuchs et al. (1977)), Moire topography (Idesawa et al. (1976)), and casting of shadow patterns (Rocker (1974)). Many of

these are attractive alternatives to stereo ranging, since the ambiguity problem is absent. However, the requirement of artificial illumination might prohibit their use in certain applications.

3.3 Shape or Structure Matching.

The relevant research focuses primarily on the comparison of 2-D projected images. An exception is the work of Falk (1970) in which matches are proposed in three dimensions and later verified by projection onto the image plane. Also, Baker (1977) compares 3-D shapes consisting of piecewise circular surface primitives, which have been built up by examining several views of an object.

Secondly, there is work which attempts to correlate a projection of a 3-D wire frame with image features. The earliest work is that of Roberts (1965). He describes a least squares optimization procedure which aligns corner features between model and image. Existence of a separate means is assumed for matching at least four points between model and image. He presents a matching technique that works for plane-faced solid objects. Hemami et al. (1975) use a hill climbing approach to solve a similar problem in outline matching. An error function relates consecutive points around an image boundary to boundary points of a model projection. Recently Barrow et al. (1977) describe another hill climbing technique in which nearest neighbor

edges between scene and model determine the error function. The authors suggest using higher level cues for better matching. Their error function resembles a fuzzy template matching technique described by Tasto and Block (1974).

Methods for comparing 2-D shapes include both symbolic techniques (Grape (1973), Perkins (1977)) and Fourier techniques (Dudani (1973)). Grape models polyhedra as a set of 2-D projections. He matches them to instances of such objects in an image by comparing 2-D angular configurations of scene edges. He allows a tolerance on observed junction angles to account for slight rotational degeneracies of the various prototype views. Another 2-D technique which includes curved edge descriptions is that of Perkins (1977). He describes a technique for interpreting occluded views of essentially 2-D machine parts using a camera oriented above an assembly line. Object shapes are represented with piecewise circular and linear segments approximating object outlines. Matching is organized so that maximum likelihood shapes are proposed first, based on a set of computed features. A match is proposed by correlating intrinsic slope functions of curves, taking into account symmetries of various shapes. It is verified by extending prearranged rays from the model to the instance. Intersection of a minimum number of rays with the instance decides the outcome. The approach seems quite general, and it succeeds even on noisy scenes with minimal information.

Dudani (1973) uses 2-D projections as models for aircraft outlines. He describes shapes with Fourier coefficients of the boundary curve for a set of orientations which covers all possible views of the body. Identification proceeds by first extracting the boundary curve of an unobstructed view of an unknown craft. He then computes its Fourier coefficients and searches for the best match to the coefficients of a view of a known aircraft.

Jarvis (1976) and Pavlidis and Ali (1977) treat the matching of linearly segmented curves by a syntactic method employing regular expressions. Problems in using this technique, though, are the specification of admissible expressions for pattern matching. This is really the segmentation problem in disguise.

3.4 Monocular Vision.

For vertex based segmentation and identification of bodies in simple scenes see the works of Guzman (1968), Waltz (1972), and recent works on extension to curved objects by Turner (1974) and Chang (1974). In this approach to visual recognition vertex projections are modeled as opposed to specific object shapes. In the progression of ideas from Guzman's body finding through Waltz's network of constraints one observes the richness of semantic information that can be obtained without specific object models. This was a significant departure from earlier

thoughts in computer vision. In a similar manner Burr and Chien (1976) show how interaction of boundary vertices (concavities) with region primitives permits body finding in occluded real scenes of irregularly shaped objects. The minimal spanning tree technique proposed by Zahn (1971) is used in organizing the visual data.

Rubin and Reddy (1977) apply results of speech understanding research (Lowerre (1976)) to a problem involving feature labeling in visual scenes. Based on a "beam search" technique, it makes use of statistical information as well as structural relationships between features (primitive picture elements) in real images.

3.5 Curve Fitting.

Turner (1974) uses a conventional least squares method for 2-D arc and ellipse fitting in pictures. Tsuji and Matsumoto (1977) show a technique for fitting ellipses in which first a center is found and then the ellipse size and orientation are computed. This is similar in style to the circular arc method presented here.

Shirai (1975) fits curves by measuring curvature as a function of arc length. Local peaks in this function indicate breakpoints. Ellipses and line segments are fitted between them.

Fletcher (personal communication) describes several techniques in his work on circular arc approximations to edge contours. In one method the enclosed area between a contour and the line connecting its endpoints is measured. There is a unique arc which encloses the same area and intersects the two endpoints. He describes a second method called the "equal r. m. s. radius" method in which an arc center is searched along the perpendicular bisector of the line segments connecting the contour endpoints. The criterion used is that the average squared distance to all the contour points from the center estimate equals the squared distance from the contour endpoints to the center. The search reduces to an analytic representation, thus making it attractive, but the endpoints of the arc are necessarily constrained to coincide with the contour endpoints. He describes also a "polar transformation" technique in which a polar coordinate system (r, θ) is affixed about one endpoint of a contour chain. The (r, θ) coordinates of the first few contour points can be represented approximately by a line segment. The theta intercept determines the angle offset between the initial tangent of the arc and the $\theta=0$ axis. The values $r/\sin\theta$, when averaged, serve as an estimate of the arc radius. All methods can be used successively to approximate an arbitrary curve based on an error criterion of fit.

Mckee and Aggarwal (1975) compute a variation of a smoothed Freeman code of a contour and fit this representation with line segments. Segments of nonzero slope correspond to circular arcs of radius 1/slope and zero-slope lines correspond to linear segments on the original curve.

Approximation of curves with splines is an area of great interest in computer aided design, but spline fitting has had little application in shape recognition studies.

3.6 Contour Smoothing.

Montanari (1970) approximates digital contours with minimum-perimeter polygons (MPP) computed by a nonlinear programming technique. This can be construed as smoothing of the curve, since the polygonal boundary is generally a better approximation to the original than the discrete data points. Sklansky et al. (1972) compute the MPP by a faster method simulating the application of a stretched string to the contour. Bennett and MacDonald (1975) show linear smoothing of a contour by truncating upper terms of Fourier series of the slope function. Mckee and Aggarwal (1975) smooth Freeman-coded contours by fixed-interval averaging.

Dynamic smoothing techniques exist in application to enhancement of grey scale images. Lev et al. (1977) show a technique for smoothing two-dimensional intensity functions with pattern-weighted averaging windows. Weights are based

on the nature of the edge passing through the window, and are uniform when none is present. A previous paper by Graham (1962) describes an edge-detection-based technique for snow removal from real time television images. Nahi and Habibi (1975) present a statistical method for dynamically smoothing a textured pattern on a smooth background. They estimate the local statistics of an image area corrupted by noise and decide whether it is the figure or background. Based upon this decision they choose one of two different filters, each optimized to the given statistics.

4. PROBLEM FORMULATION

Throughout this work computer vision is structured as a two-level process. Taking such an approach, one generally assumes that image data can be organized symbolically based on edge and/or region descriptors, independent of the application domain. The description is a reduced one, containing essentials for recognition, but is otherwise fairly complete. The output of this stage serves as a "description" of the scene. A similar "description" of objects in the domain is embodied in a model data base. The second-level process attempts to compare scene and model "descriptions" to arrive at a higher level domain-dependent "description" of the scene, based on instances of object models.

Past efforts in computer vision have taken a similar two-level approach. However, results have generally been plagued by the "hourglass problem". That is, much is known about scene segmentation and low level processes in addition to high level or semantic processes. However, the link between the two has not generally been studied in regard to its possible influence from design constraints at each level. Therefore, little is yet known about inter-level interaction. The depth extraction approach is taken here as one attempt to remedy the situation, in anticipation that similarities in the descriptions of model and scene might

permit more straightforward interaction.

The problem of 3-D feature extraction from picture pairs can be approached by first constructing a depth map of the scene, and following it by segmentation of the implied 2-D surface. In some cases this may be the desired approach, particularly in cases where objects are modeled by surface descriptions or medial axes. Unfortunately this usually requires high information density in the picture or pronounced textures on object surfaces. The focus of this work is on modeling man-made objects as opposed to natural or outdoor scenes. Though such surfaces are often smooth, they usually have prominent edges, so an edge-based approach would seem more fruitful.

Edge patterns can be modeled in a number of ways ranging from polynomial functions and Fourier representations to discrete, or syntactic, descriptions. The discrete approaches are favored for computer vision since they permit straightforward handling of occlusions and incomplete descriptions. Two popular discrete methods for 3-D contours are piecewise linear and piecewise circular wire frames. With linear representations much efficiency can be gained by limiting depth computation to breakpoints or line ends. Bulk correlation techniques normally require that the pattern window be taken at a place of high intensity variance, so that sharp peaks can be found in correlation search. Focusing depth computation to the

vicinity of strong intensity edges satisfies this requirement automatically. Further problems are encountered with simple edges due to the fact that such variance is directional (see Figure 4-1).

Object modeling is an important concern in machine perception, since domain-dependent interpretation is often required in practice. Computer graphics research is naturally concerned with geometric modeling of objects, and thus a wealth of information exists on the subject. If scene edges are described with linear segments, then a similar description of object models would allow straightforward model matching techniques. Modeling of scenes and objects with circular segments, though, is attractive for reducing the verbosity of shape descriptions and thus search combinatorics. Since circular segments do not preserve circularity upon central projection, attempts to describe shapes with circular arcs might be facilitated if segmentation is done in 3-D, after an edge depth map is obtained. Since this would generally require more extended search, the matching technique should be efficient. One such approach is to arrange the two views so that differences are minimal (narrow angle approach), thus permitting simple and rapid techniques for feature comparison.

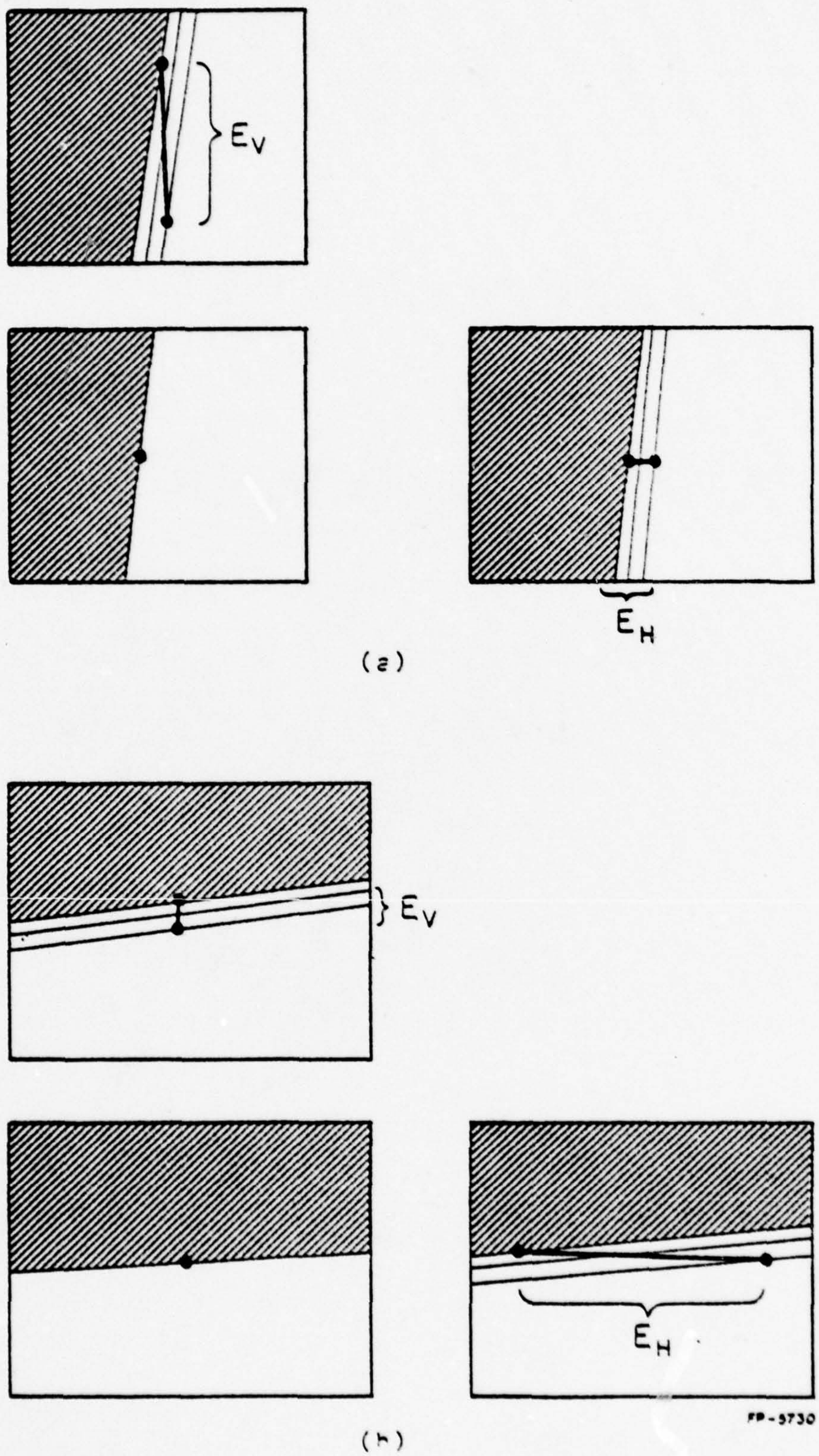


Figure 4-1 (a) Illustration of matching error for a vertical edge segment relative to vertical and horizontal shifts of the scene. Small errors in position due to field geometry or camera orientation cause appreciable matching error in one view. (b) Same for horizontal edge segment.

The narrow angle approach introduces a serious problem in triangulation accuracy. Namely, due to picture digitization and noise, the depth accuracy may be quite poor. Nevatia (1976) gets around the problem by tracking features incrementally over a sequence of image rotations, saving ray intersection until a large angle is reached. This technique is successful, but in certain applications the need to rotate the scene or camera might prohibit its use.

Reconstruction of 3-D scenes is only the first step in the stereo vision process. Since these structures invariably contain errorful, missing, and occluded edges, robust model matching techniques are essential for their interpretation. Stereo correlation can be successfully implemented using relatively local information, since camera models, and depth continuity can be employed to effectively constrain search. In addition, both views are known a priori to contain a projection of the same scene. However, in matching scenes to a set of object prototypes (model matching), much less information is available, and thus more extended or global edge descriptors should be used so that search combinatorics remain manageable. This is the reason for edge segmentation with lines and arcs. Means for determining validity of proposed matches needs to be incorporated, so that decisions can be made to control the matching process. Matching efficiency in cluttered scenes and with large data bases would also be topics of concern.

Although a variant of 3-D template matching might be useful for recognition of rigid shapes, more powerful techniques are needed when objects are flexible or classes of shapes must be understood. Much research must yet be done on the specification of admissible distortions for particular domains of application.

Several problems associated with obtaining 3-D features from stereo image pairs have been formulated. They consist of locating and organizing edges into symbolic descriptions, stereo matching of picture features, and minimizing errors associated with edge directionality. For computation of edge depth maps there are added problems of matching efficiency and continuity implementation, as well as the symbolic representation of contours. In relation to matching of 3-D structures several problems are encountered -- matching strategies, representation of shape, occlusion, spurious and missing edges, and search efficiency in complex scenes and in situations with many models.

In Chapters 5 through 8 solutions to the above problems are presented, and results are discussed in Chapter 9.

5. METHOD OF MULTIPLE VIEWS

Approaches which require high level features for stereo matching (vertices, connectivities, etc.) usually perform less than adequately since these features are often incomplete or hard to define for complex real scenes. Also, little theoretical knowledge is yet available on the matching of nonisomorphic graphs. A simple template matching technique known as bulk correlation has proved reliable in stereo comparison, but the objects must usually contain texture or surface detail. The approach used here is to modify bulk correlation methods for use on scenes containing smooth arbitrarily shaped objects. *Redundancy* provided by several views is used to enhance the correlation peak at low information windows. It presupposes that viewing parameters are known for all views. Use of multiple views here differs from that of Rabinowitz (1971) and Shapira (1977) in that no restrictive assumptions are made about surface shape. In addition, bulk correlation techniques are used to enhance the voting between views.

In this chapter and also in Chapters 6 and 7, techniques for computing central projections and their inverses are implied. Since such techniques are standard and available elsewhere, their details have been purposely left out of this work. For details the reader is referred to such references as Roberts (1965), Duda and Hart (1973),

and Hannah (1974). Techniques for measuring camera focal length and aspect ratio are taken from Baumgart (1974).

5.1 Redundancy Effect.

From a feature location in an image, it is possible to simulate the projected ray passing through it, provided the constituent parameters (i.e. camera orientation, position, and focal length) are known (Hannah (1974)). The particular feature observed in the image could have originated at any 3-D location on this projected ray. This is precisely the problem of ambiguity. Stereo comparison methods attempt to resolve this ambiguity by viewing the same scene from two or more different aspects. The idea is to specify an additional ray for the feature, thus defining its location by ray intersection. This requires capabilities for comparing features between images. A feature which has few distinguishing characteristics (i.e. edge segments) may in fact match well to a number of locations in the other view. Even when one simulates the ray from the original picture and projects it into the second view, there may still be several ambiguities. In general, the less discriminating is the feature, the greater the chance of ambiguity in the other view.

In Figure 5.1-1a we see an illustration of the redundancy effect for a two-dimensional object. The vertices of the object are numbered from 1 to 7. The

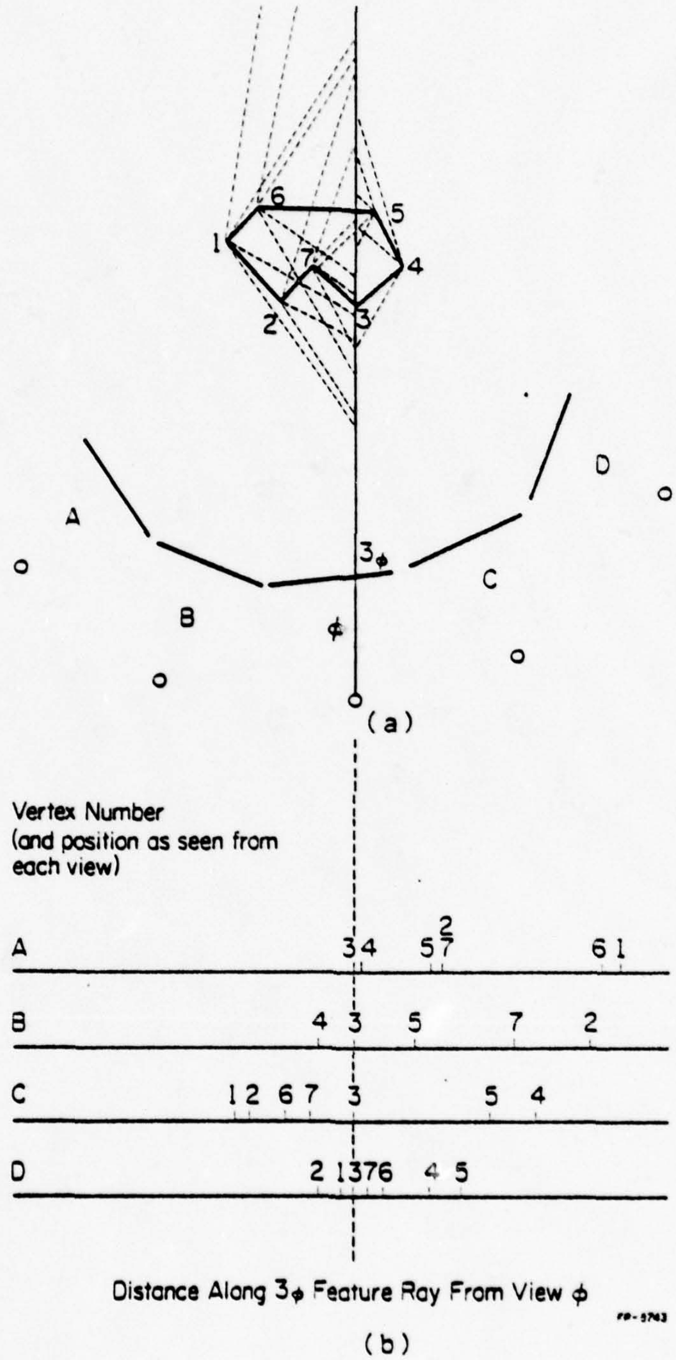


Figure 5.1-1 (a) Illustration of multiple view redundancy in two dimensions. A feature from vertex 3 in the object appears at location 3_ϕ in view ϕ. Additional views A, B, C, and D are shown, and corresponding features of the object are projected from these projection centers onto the ray 3-3_ϕ. (b) The locations of projected features along line 3-3_ϕ of 'a' are shown. Notice that the projections of the true feature (number 3) all occur at the same ray coordinate, and others are misaligned.

symbols A, B, C, D, and ∅ indicate various views and projection centers. The number 3 feature projects into view ∅ at location 3∅. Observe the various intersections of other feature rays (dotted) with ray 3-3∅ through projection centers A, B, C, D. The respective ray locations of the intersections are indicated in Figure 5.1-1b for each view. Notice that object features positioned off ray 3-3∅ project onto ray 3-3∅ at different locations. This means that to a certain extent depth may be estimated by mere voting, requiring no discrimination between feature types (i.e. edges or corners). While stepping along the feature ray 3-3∅, one can project rays simultaneously into all images, recording the number of views having features at the projected locations. The ray coordinate with the highest count is taken as the depth for that feature.

When polyhedral vertices are used as features with four or so views (Rabinowitz (1971)), this approach is adequate and results in reliable stereo matches. However, for local and less discriminating features such as simple edge elements, either more views must be used or the voting must be enhanced by an additional means. Shapira (1977) adds the property of vertex ordering and is able to extend the domain to polyhedra with simple curved surfaces. For more complex shapes with arbitrarily curved edges, the technique is better enhanced by cross correlation. (Enhancement by symbolic edge properties is inadequate since often comparisons must be made at corners as well, where simple

edge properties are inadequate. Cross correlation is more general, treating edges, corners, and arbitrary feature patterns.) Cross correlation does not require prior segmentation, thus bypassing difficulties normally encountered in comparing irregular segmentations. Because of the potential inefficiency of bulk techniques, though, they should be used sparingly. Observe in Figure 5.1-2 three examples showing the result of summing cross correlations (mean square difference measures) from two pairs of stereo views. The prominent peak is enhanced in each example.

5.2 Edge Tracking and Contour Approximation.

The use of several views is tied to a scheme for efficiently conducting the search and restricting depth computation. In this method three pictures are required, a center, north (10 degrees), and east (20 degrees) picture. The center picture is preprocessed to locate edges using a modified gradient operator (see Appendix) which searches for local gradient extrema. The edge features are passed to an output list, retaining x-y location, direction, and gradient magnitude. Edges are tracked in near neighbor fashion by a circular scan which first tests nearest neighbors and then others until a search radius of three pixels fails to indicate an edge. When the tracker fails, left to right scan is resumed in search of another edge contour.

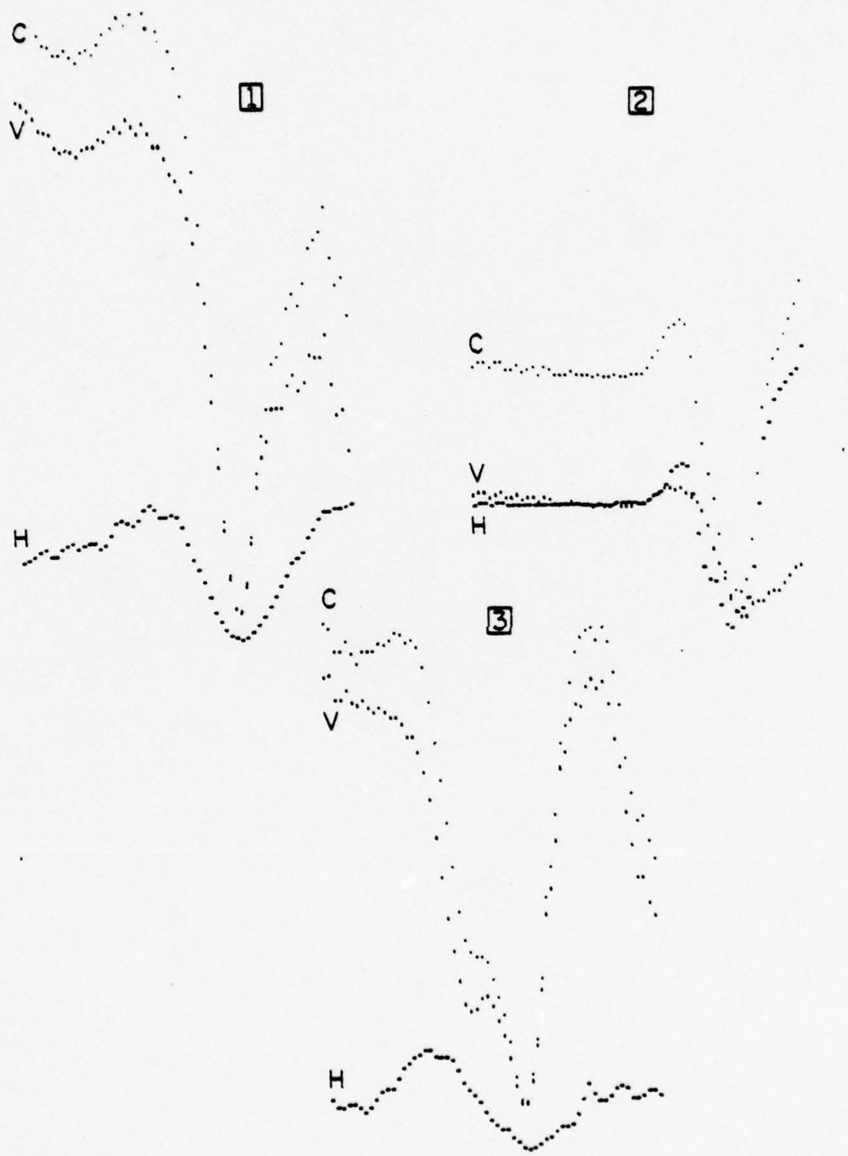


Figure 5.1-2 Examples of multiple view redundancy in enhancing the correct correlation peak. The three figures correspond to three correlation points taken from Figure 9.2-2. Correlation value is plotted against depth for each of horizontal (H) and vertical (V) picture pairs (C=V+H).

The edge list is passed to a contour approximation routine which describes each contour with an end-connected sequence of line segments. The method used is the iterative endpoint fit as described in Ramer (1972) and in Duda and Hart (1973). In this method successive approximations are made to sections of the contour in a recursive manner. The first approximation used is a single line connecting the endpoints of the contour. Perpendicular distances from all contour points are then measured to this line. The point having the greatest distance serves as a new breakpoint, if it exceeds a threshold, and two new segments are formed as a second approximation (see Figure 5.2-1). This procedure is recursively called on each new segment formed, splitting the previous approximation, until a condition is achieved such that all contour points lie within a preset tolerance of the segmented approximation.

Other linear approximation techniques can be used but this was found adequate for the purpose intended, even though segment ends are necessarily constrained to lie at discrete points. A variant is to use this as a first approximation and improve the fit of each segment by least squares (see Section 5.6). Pavlidis and Horowitz (1974) merge an initial approximate splitting according to a merging criterion. Horowitz and Pavlidis (1974) apply the split and merge technique to region segmentation. If the scenes are very noisy and edges are hard to track, then a more global line finding technique such as the Hough variant

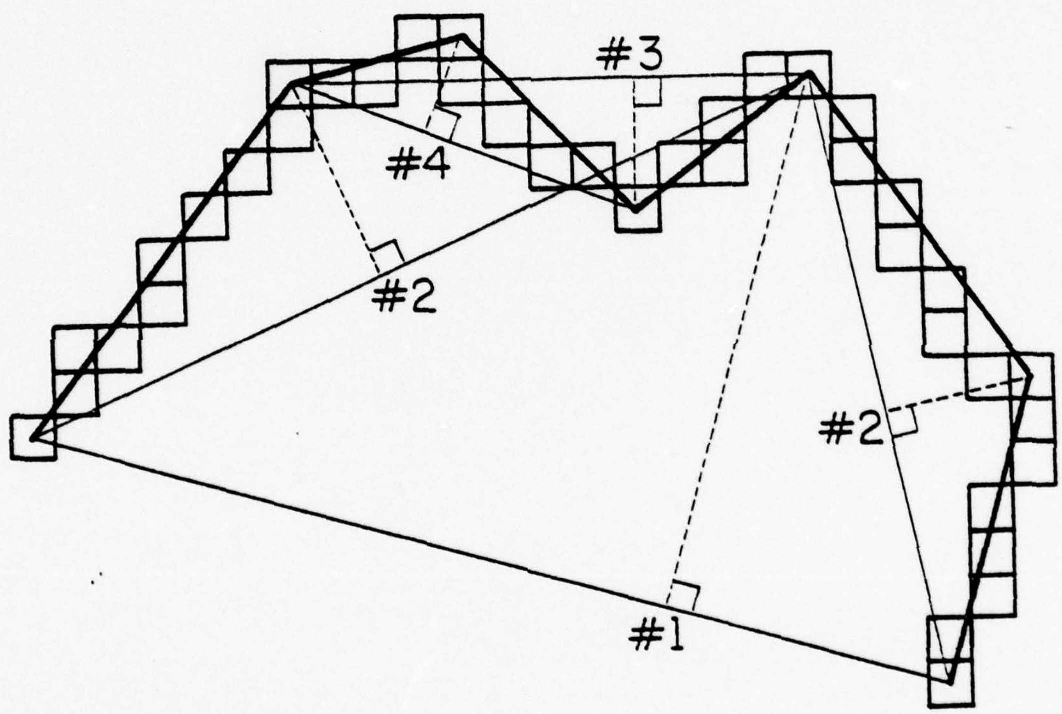


Figure 5.2-1 Illustration of the iterative endpoint contour approximation technique of Ramer (1972). The successive levels of splitting are indicated by numbers 1 through 4, and the solid lines indicate the final approximation.

of Duda and Hart (1972) might be used. However, it is more costly, since it effectively searches all edge candidates in a picture for inclusion in a line.

5.3 Correlation Matching.

Small length segments in the edge list can be filtered out prior to stereo matching, since they would contribute less to the 3-D structure than long segments, for the same amount of search computation (or they can be passed to a merging program which links edge segments with proper orientation and proximity). The filtered list is then passed to the multiple view correlator, which attempts to find a 3-D coordinate for each segment junction. Correlation can be implemented as a product or a difference, the latter being cheaper but not as general. In scenes where the illumination is approximately the same in each view, the difference method is adequate, whereas the product normalizes out intensity and contrast differences between pictures. Nevatia's form of the mean square difference operator is used (Nevatia (1976)).

$$\text{M.S.D.} = \frac{\sum_{ij} \left(P1(i,j) - P2(i,j) \right)^2}{\left(\sum_{ij} P1(i,j) \right) * \left(\sum_{ij} P2(i,j) \right)} \qquad (1)$$

P1(i,j) and P2(i,j) represent the intensity arrays of the two pictures, where indices (i,j) are relative to test

centers in each image, and summations are over rectangular windows. Windows of size 9 X 9 were most often used as patterns, located at segment breakpoints in the center picture.

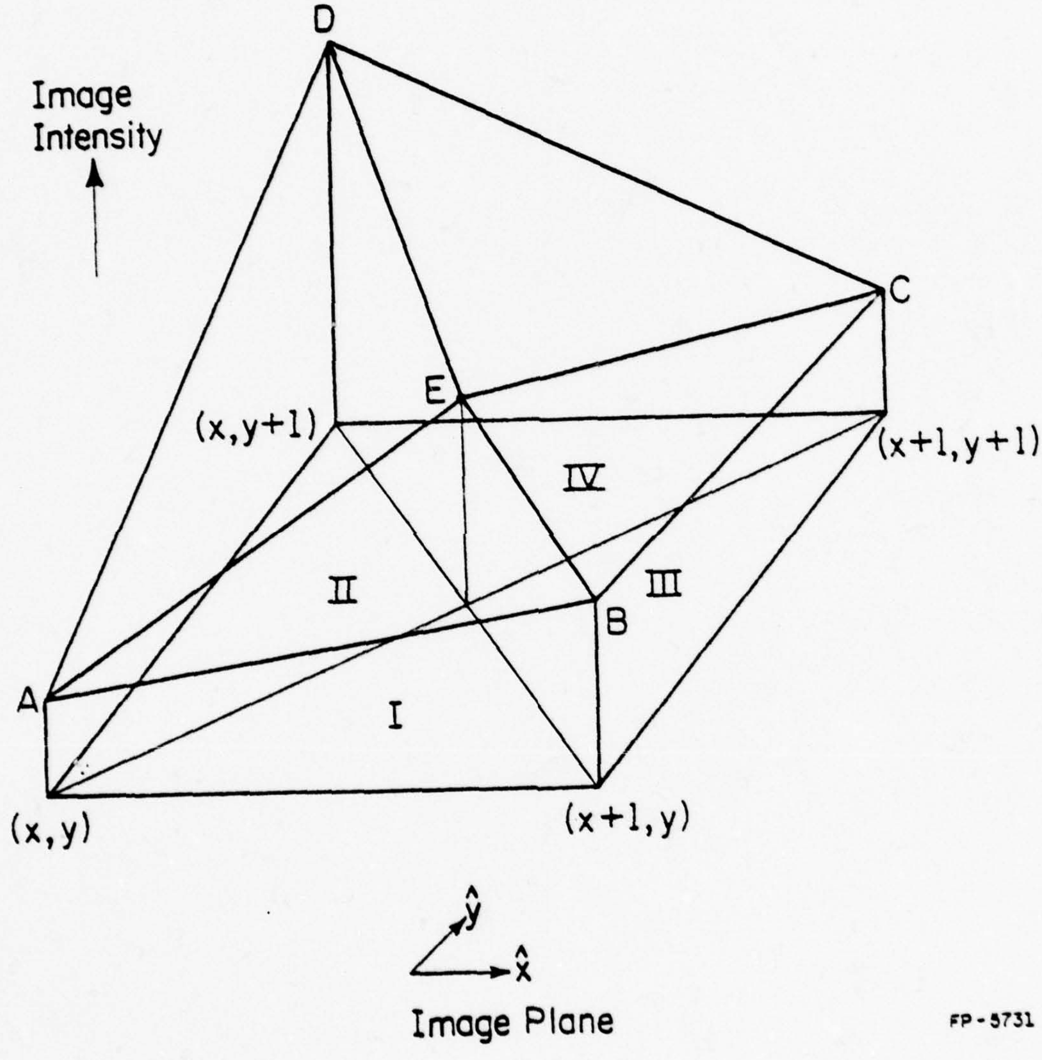
Search is performed in the north and east picture simultaneously by simulating the projected ray from the center picture through the given feature location, and then back projecting successive points along this ray into the north and east pictures respectively. A new composite correlation is defined as the sum of the two pairwise differences (Equation 1), using center-north and center-east picture pairs. Knowledge that objects are expected to lie between two depth extremes restricts search to only a small segment of this line, which can be easily computed. Furthermore, if a cheaper operator can be used to prefilter test points, then search can be further restricted. Since all features lie on object edges, filtering is accomplished by first computing an edge operator, the same one used in the center picture, to eliminate candidates which have insufficient edge strength to match the pattern. A fixed threshold is used, slightly less than the one used in the center picture. This affords an order of magnitude or more savings in time, since only ten or fewer pixels are addressed in the edge operator, whereas roughly 100 are addressed in the correlator. For another approach to search reduction for registration of images see Barnea and Silverman (1972).

Since 3-D points along the search ray will generally project into noninteger values in each picture, a means of interpolating the mean square difference operator had to be added. The interpolation algorithm utilizes piecewise planar approximation of the picture intensity function (see Figure 5.3-1).

The global minimum of the operator over the range of search is taken as the correct match, and its location is improved by parabolic interpolation between itself and adjacent neighbors (± one step). Roughly 60 depth steps are used in the search for a typical scene. Since the test points are projected from a 3-D location, the 3-D coordinate is already known, requiring no triangulation.

5.4 Two Variations.

Two alternate techniques were tried in conjunction with bulk correlation as defined above. One attempts to compute depth points incrementally along edge contours of the central picture, by using the difference operator at each edge element location. Search for the first point of a chain proceeds as described above. Successive points however are constrained to match in a small interval (± 4 steps) about the last computed depth value, so as to preserve 3-D continuity. This was found less desirable than the narrow angle method (Chapter 7), which also enforces 2-D continuity of edge chains in each picture.



FP-5731

Figure 5.3-1 Illustration of planar interpolation of intensities between discrete pixel locations. If E is defined as the average of the intensities A, B, C, and D, then four planes are specified, ABE, BCE, CDE, and DAE. The interpolated intensity value for an arbitrary location within the square region becomes the value indicated by the planar segment directly above it. This results in a continuous intensity surface with piecewise constant gradients.

Another technique uses only pairwise correlation as opposed to composite correlation, but restricts search to either a north or an east picture, depending upon whether the local edge feature is respectively horizontal ($\leq 45^{o}$) or vertical ($> 45^{o}$). See Burr and Chien (1977) for this approach.

5.5 Band Search.

Other techniques for searching for intensity templates allow search in a band centered about the projected feature ray. This is necessary when searching for textured features since geometric distortions in the viewing field cause the correct feature location to appear off the search line. If search is restricted to lie only on the line, then the correlation peak in the vicinity of the correct feature would be shallower, and might be confused by other peaks. When correlating edge features, however, band search is not necessary, provided that two orthogonal view pairs are used as described. Geometric distortion still dislocates the feature, but its counterpart on the search line will be less perturbed provided the edge intersects the search line roughly orthogonally. Thus the counterpart serves as a good alternate match. In addition, the match error is primarily perpendicular to the search line, and ray intersection is thus mildly affected. Ray intersection errors are affected primarily by match errors parallel to the search line. Therefore when three pictures are used in the scheme as

described, the primary contribution to depth error is from geometric distortion along the search line. Matching error is minimized by having two orthogonal view pairs.

In addition, linear as opposed to band search solves the sliding vertex problem (Pingle and Thomas (1975)) quite elegantly, since a sliding vertex normally moves away from the search line, and thus does not confuse the matcher. A sliding vertex is a false vertex (tee joint) arising from two edges at different depths, which appear to intersect as viewed.

5.6 Refinement of Edge Approximations.

Hill climbing techniques can be used to improve contour segmentations. Ramer's (1972) method results in line segments whose endpoints lie on contour elements. Thus the approximation may not be optimal in the least squares sense. However, the solution is close enough so that iterative feedback can be used to improve it. Based on proximity of curve points (xi,yi) to the initial segment approximation (those within a delta neighborhood of each segment), a standard least squares error function is defined for each segment as follows:

$$\mathrm{ERR} = \frac{\left(\sum yi^2 + C1^2 \sum xi^2 + n * C0^2 + 2 * \left(C1 * C0 \sum xi - C0 \sum yi - C1 \sum xi*yi \right) \right)}{\left(1 + C1^2 \right)} \qquad (2)$$

where C1=((y2-y1)/(x2-x1), C0=y1-C1*x1, and the pairs (x1,y1) and (x2,y2) are the endpoints of the proposed line segment. Summations are performed over all points in the delta neighborhood. The square error rather than mean square error is used since more data points on a segment should bias the fit toward that segment. Also, the error for a particular breakpoint contains two error components, one from the segment on each side. Breakpoints along the contour are iteratively perturbed, and the error function is locally minimized by following steepest descent paths (see Figure 5.6-1). When the condition is reached such that all breakpoints are at local minima, then the process is terminated. Though the error is local, the iterative nature of the breakpoint adjustment propagates information throughout the contour, so that local constraints get influenced by global ones.

The same idea can be used for improving 3-D coordinates of structures computed using multiple views (chapter 5). Since the bulk correlation measure for matching images is necessarily local, the resulting match may not be globally optimal. Upon observing the projection of each structure into the original stereo images, one sees that most corners are reasonably close to alignment, but edge fits could be improved. By redefining the error function at a 3-D breakpoint as the sum of each 2-D <u>projected</u> error (Equation 2), the hill climbing technique can be extended to 3-D structures. The constraint that a unique 3-D intersection

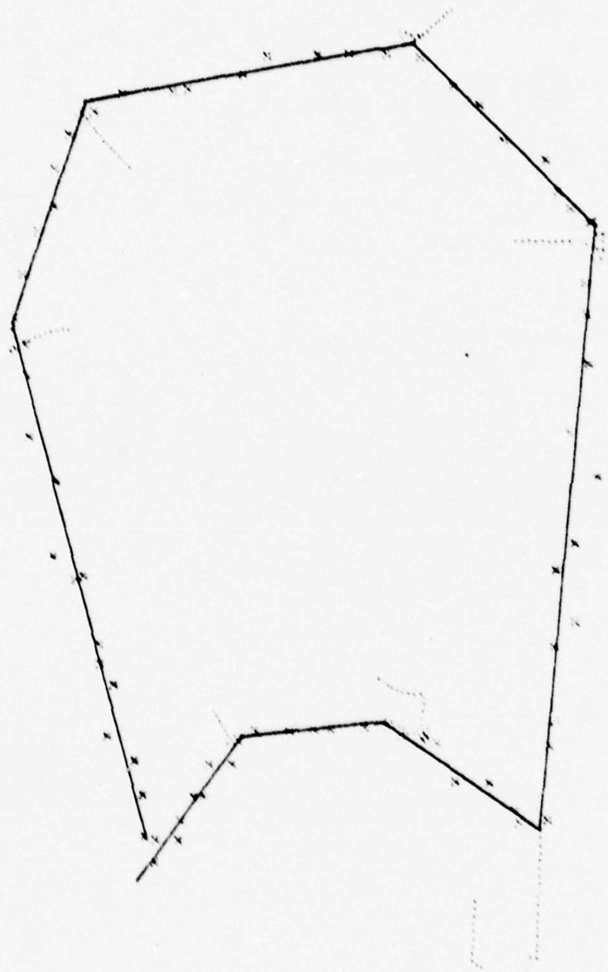


Figure 5.6-1 Illustration of hill climbing technique for improving contour approximations. The paths located at each breakpoint indicate loci followed during algorithm iteration. The final segmentation shown is the one satisfying square error minimization at all breakpoints.

point exists is enforced by choosing the 3-D point first and then projecting. By successively perturbing the three spatial coordinates (x,y,z) at each breakpoint, the error is minimized until all breakpoints are at local minima, similar to the 2-D case.

6. MODEL MATCHING

6.1 Wire Frame Modeling.

Except in some degenerate cases, projection of a 2-D line segmentation into three dimensions results in a structure very similar to that which one would obtain by approximating the actual 3-D edges with linear segments. This in fact is the representation used for an object prototype, namely a wire frame exoskeleton in which linear wire frame segments approximate object edges or loci of high surface curvature. An example of a degenerate case is a planar curve projecting into a single line.

When sharp edges do not exist, loci of relatively high curvature may be substituted, since they would be most likely to predominate on object silhouettes. Most man-made objects, especially machine parts, are represented well by their prominent edges. When objects have no characteristic edges then some edge network may be substituted (e.g. a dodecahedron for sphere). The modeling scheme is admittedly biased toward objects with prominent edges, but will work, though less adequately, for edge-less objects.

In modeling of any sort, one attempts to describe observable features, since at some point comparisons will be made to the real world. Wire frame structures can simulate properties such as surface orientation, hidden lines, shading, as well as edge structures. However, in the

absence of shading techniques (Horn (1970)) and illumination information, 3-D surface orientation is not readily obtainable. Therefore, the choice was to exclude surface information. In addition, since hidden line elimination can be quite costly, it was felt best to develop techniques which do not require it. As a result prototypes consist only of information about node position (corners, high curvature points) and their connecting structure (wire frame). Addition of a hidden line algorithm would only reduce the number of model edges that need to be compared at any given time, and it is not clear that the increased computation cost would be offset by increased matching efficiency. This would be an interesting topic for investigation.

In the remainder of this thesis the terms "percept" and "3-D scene" will be used interchangeably to refer to 3-D edge constellations produced by the multiple view process (Chapter 5).

6.2 Matching 3-D Wire Frames.

The matching process attempts to pair percept and model edge descriptors on the basis of their relative lengths, orientations, and positions. The process relies on constraints imposed by 3-D geometry to rule out impossible relationships. It is similar to the approach used by Falk (1970) in that two steps are involved, a proposer, and a

verifier. It differs from his method in that matching takes place entirely in three dimensions, not requiring projection into the image to verify a proposal. It has the flavor of template matching in that rigid structures are compared, but the nature of the matching is symbolic. The ideas are compatible with subtemplate strategies of Vanderbrug and Rosenfeld (1977).

6.2.1 Proposer.

The proposer attempts to relate a single pair of edges in the percept with a pair in the model. In the process the constraints shown in Figure 6.2.1-1 must be satisfied. Essentially they require that lengths must agree, as well as the position and orientation of the two edges relative to each other. As implemented the routine first selects arbitrarily an edge segment of the percept ($\overrightarrow{VP1}$). A model edge is sought whose length ($|\overrightarrow{VM1}|$) nearly equals $|\overrightarrow{VP1}|$. When such is found, there is still a mating ambiguity: head to head, or head to tail. First one is tried, then the other. The two edges become respectively the z-axes of cylindrical coordinate systems centered at each edge. The rotational ambiguity is resolved by searching for an additional edge in the percept which matches a model edge on the basis of similar lengths, cylindrical (r,z) coordinates, and orientations, to within a fixed tolerance. If and when such is found, the pairing defines a coordinate transformation between model and percept, namely that which

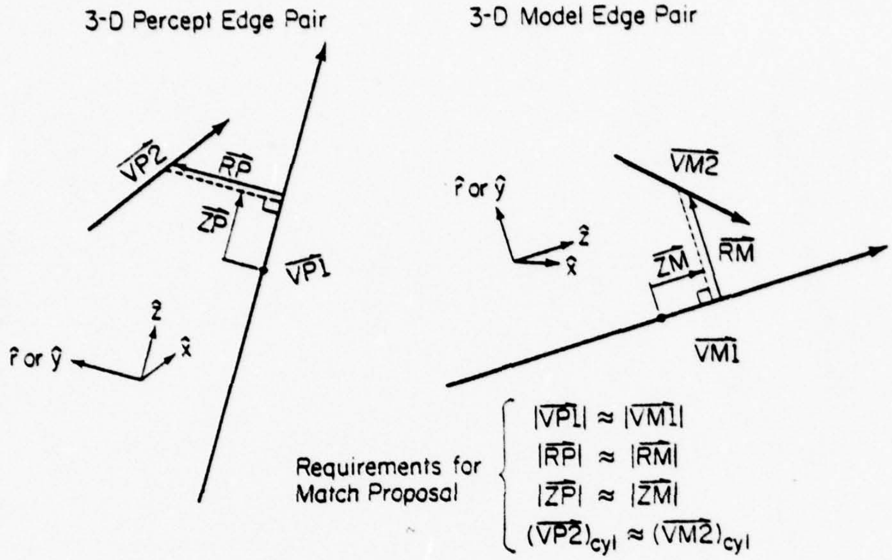


Figure 6.2.1-1 Illustration of the constraints required for proposing a match between the scene (left) and the model (right). A correspondence between two scene edges and two model edges (heavy lines) is required. Essentially the entire structures must agree geometrically, within error bounds.

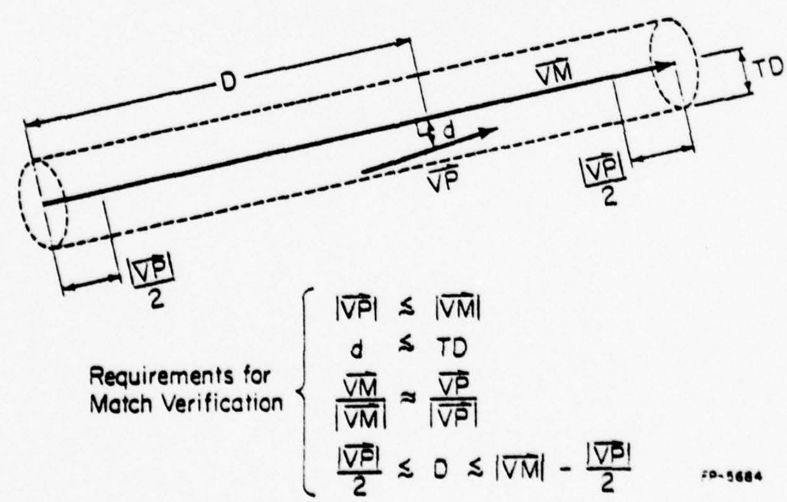


Figure 6.2.2-1 As part of the verification of the proposed coordinate transformation, remaining implied matches between scene and model are checked. If the above constraints are satisfied, then the CONFIDENCE value of the proposed transformation is increased. In order that scene edges with missing segments may contribute to match validity, strict agreement of edge lengths is not enforced. VP and VM correspond to percept and model edge vectors, respectively.

brings the three points into correspondence: two ends of segment 1, and the center point of segment 2. There may exist further implied edge matches between the two structures, so an attempt is made to verify the goodness of the complete structural match by finding these implied matches. This is referred to as the verification step.

6.2.2 Verifier.

In verification it is more natural to implement comparisons in a cartesian rather than a cylindrical reference frame. These are defined as shown in Figure 6.2.1-1. In fact, the implied rotation is not actually performed, but instead all that is needed is to compute edge information relative to each cartesian frame and compare coordinates. It is desirable because of possible missing edges and/or segmentation uncertainty, to allow a portion of a percept edge to match a model edge, and relax strict length agreement. This could also be allowed in the proposer, but was not, since search time would be increased. At any rate, it would be a simple extension to allow this in a practical system, or an alternate view can sometimes be taken in practice. It is usually possible to find at least two complete edge segments present in an object unless the scenes are extremely noisy. In the process of searching for a proposal pair, all matings of scene edges are tried, since it cannot be assumed that any given pair has a correspondence in the given model.

In verification the constraints shown in Figure 6.2.2-1 are enforced. The essence is that a percept segment must agree in orientation with the model segment, and its position must be effectively within a cylindrical shell about the model edge, but its length may be less than that of the model edge. The tests are ordered in an attempt to make the search somewhat efficient. Absolute features such as edge lengths are tested first and are precompiled. Relative features which cannot be compiled, such as distances and angles, are tested later.

In verification all pairings of percept and model edges are searched for this criterion. When one such pairing satisfies the constraints, it is taken into account to increase confidence in the match. This is done by computing a running sum, called the CONFIDENCE. This value contains the normalized sum of all percept edge lengths which match at least one model edge. The normalization factor is just the sum of all percept edge lengths. Thus a validity measure is assigned to each proposed coordinate transformation between a particular model and the 3-D scene. This number allows us to make comparisons between the goodness of fit of several orientations of one model, and between several models. Search can continue over the data base testing various models and orientations until one is found which maximizes the CONFIDENCE value. If the value exceeds a threshold (typically 0.5 or more), it can be taken as the correct identification of that portion of the scene.

The matched lines are eliminated, and the remainder of the scene is searched for further matches. Efficiencies can be gained by restricting such matching to subsections of the scene, and to subparts of the models (Section 6.4). Examples are shown for matching scenes consisting of single and multiple objects. Extensions to the linear modeling scheme include one based on circular primitives.

6.3 Edge Connectivity.

The reader will observe that edge connectivity information was not utilized in the matching process to aid in search reduction. Although it would be desirable to implement this, it is not practical to do so in general. Often a continuous edge is broken due to illumination or noise, and there may be depth errors of great magnitude at isolated points. An alternative is to enforce connectivity until a mismatch occurs, or a contour ends, and then allow unrestricted search until a new match is found. This approach would necessitate some means to deal with the problem that the correct pair may be segmented differently. The problem is treated here by not enforcing connectivity, at the expense of increased search. Errorful coordinate transformations are often proposed when the proposal pair is restricted to near neighbor edges (due to uncertainties), especially when located at a low curvature point. This would need to be counteracted with some fancier verification stage, perhaps fuzzy near neighbor matching (Barrow et al.

(1977)) rather than template cutoff past a certain range. Alternatively, an incremental update on the coordinate transformation might be tried as new match points are added. Nevertheless, the chamfer matching technique of Barrow et al. (1977) requires hidden line elimination for optimal performance. It is not clear that the tradeoff of more complex verification and costly line elimination would make this approach better. Perhaps a restricted hidden line algorithm that removed only some of the edges would suffice, since some self-occluded edges are cheaper to eliminate than others.

6.4 Some Thoughts on Search Reduction for Cluttered Scenes and Many Models.

It should be clear that the shape matching routine is capable of comparing an object which is partly occluded, or one which has additional clutter surrounding it. This is due to piecewise description of shape and rich use of geometric constraints. Thus an approach to matching cluttered scenes would be to apply the recognizer to the whole scene, attempting to find a match in some region. Though this strategy works, it is quite inefficient, and consumes much time in checking parts of the scene which do not contain the object. In addition, CONFIDENCE values would all be low, since only a small portion of the scene would likely match at any given time.

Any prior knowledge available which permits rejecting of certain features during comparison would necessarily result in search reduction, since the combinatorial possibilities including the rejected feature are limited. This applies to both the features in the scene (3-D configurations in this case) and features in the model data base. In the case of the models, elimination of certain unlikely models or external methods for ordering likely models for testing would result in search reduction, especially if reliable knowledge is available to decide when to terminate a search.

6.4.1 Search Localization and Relaxation Labeling.

A particular technique which holds considerable promise in this regard is that of relaxation labeling (RL). It was initially demonstrated by Waltz (1972) for semantic labeling of line drawings, and recently formalized by Rosenfeld et al. (1976) to permit fuzzy constraints and labels. The technique is generally applicable when a set of objects a can take on any one of a set of labels λ_j. Labels are not assigned to objects a priori, but are assigned probabilities for each object. Probabilities are updated based on semantic constraints applicable to the particular problem domain, and particular works attempt to specify these relationships (Zucker (1977)). Semantics of label interaction between hierarchies are also important since labels can be assigned at different levels (Rosenfeld and

Davis (1977)), though hierarchical ideas are not as easily incorporated in such schemes due to their inherent nonhomogeneous nature.

Though Waltz's work indicated applicability of RL to general object features (shadows, cracks, concave, convex, occluding edges, etc.), such ideas can be extended to allow labeling of object-specific and geometric features. A simple demonstration of this is the work of Davis (1976) in which corners of a square are labeled iteratively. For general visual scenes, though, orientation-independent shape features would have to be established. For example, such features might include such measures as distance histograms between high curvature points (object vertices), or length/angle/length measures for object corners. Certain deficiencies of RL are appreciated, though, and care would need to ba taken to prevent instabilities which can occur when feedback is present. Instabilities, though, do occur in human vision (Attneave (1971)), thus lending credence to the possibility that they may be inherent in perception.

The importance of RL in geometric model matching would be to permit focus of matching to isolated regions of a scene, to order likely objects for testing at any given time, and to permit improved knowledge about match validity. During the updating of scene labels, hopefully some labels would dominate over others in portions of a scene, and the high probability labels would indicate likely objects for

geometric testing. If spatial continuity is invoked properly, high probability labels for certain subobject features would be reinforced locally by other high probability labels for the same object. Thus we would have a sophisticated object-specific body finder, or localizer, which could be used to prune or order search in both the scene and the model data base. Pruning would result since the CONFIDENCE denominator can be redefined to include weighted lengths of scene edges (varying with each test object). The measure would thus be more accurate for particular objects. CONFIDENCE values close to 1.0 would occur much earlier in the search, and this would prompt early termination.

6.4.2 Medial Axes, Hierarchical Decomposition, and Body Finding.

Alternate methods of reducing search include medial axis representations, hierarchical decomposition of models, and body finding.

The medial axis representation (Agin (1972), Marr and Nishihara (1976)) can be used to advantage in matching, since representation of shape is reduced to essential features, and thus the combinatorics of matching are reduced considerably. In fact one might consider medial axis representations to propose matches, followed by a verifier which looks at finer details, such as edge patterns.

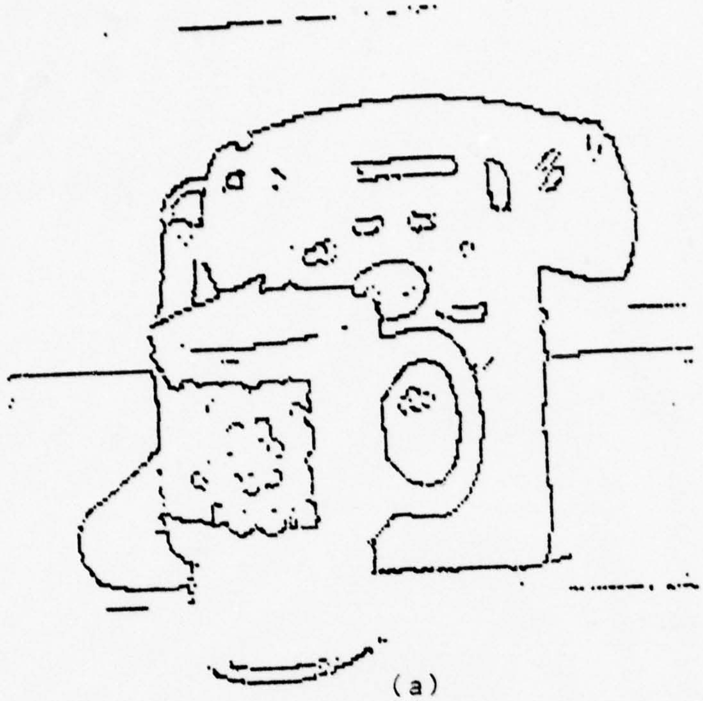
Hierarchical decomposition (Turner (1974), Schnier (1977)) is attractive for the same reasons. Namely, by searching for a subelement of a particular object, one needs to make comparisons only between a reduced subset of edge features, thus reducing combinatorics. Location of a possible subelement of an object can be followed by testing for the entire object. If the particular subelements are chosen properly, (i.e. so that an element is general and can be used as a building block for many different objects), then search is further reduced. This is possible because searching for a general subelement is essentially searching for that element in all objects where it is present, and it only needs to be matched once. This technique has indeed been used successfully in the HARPY speech understanding system (Lowerre (1976)). The presence of a subfeature then points to the objects which contain the particular feature. There is danger in using decomposition in scenes with missing information, namely there may often be insufficient information to verify a subobject. Complete models have been used until better data can be achieved. In some sense decomposition is automatically incorporated by choosing feature primitives such as linear or circular segments, as has been done in this thesis.

Body finding, although a highly developed technique for perfect drawings of scenes, is not yet sophisticated enough in real scenes, though there is no theoretical reason yet found to prevent its being used. Finding of bodies from

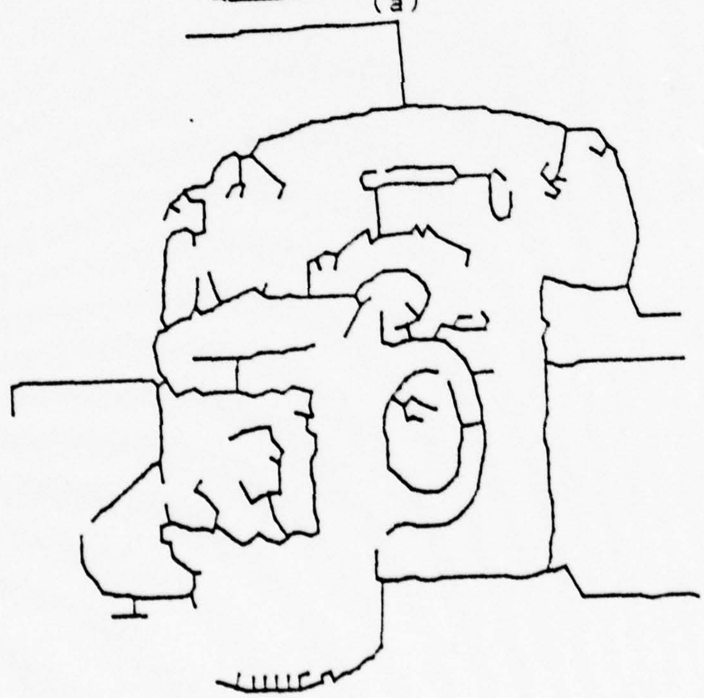
monocular data has been discussed in Guzman (1968), Falk (1970) and Chang (1974) and in depth data by Agin (1972) and Nevatia (1974). This is certainly a very strong heuristic to use in model matching, because one need perform matching only on the subpart of the scene containing the body. The minimal spanning tree representation (Figure 6.4.1-1), coupled with a generalized dynamic smoother (for 3-way and greater vertices), might prove useful for edge-based body finding. However, its success would be based on the ability to reliably detect tee junctions, and this would be enhanced by proper choice of a smoothing algorithm.

6.4.3 General Heuristics.

Some heuristics which are directly applicable in this approach are as follows (those incorporated in the current scheme are starred *): Perform matching only on connected edge elements of the scene at any given time. This resembles body finding since connected edge elements are likely to belong to the same object, especially if one stops at locations of depth discontinuity. Another heuristic is to restrict matching only to those line segments which are relatively long, since they contribute the most incremental amounts to the CONFIDENCE per element, and orientations are more accurate than short segments *. Verification can be allowed to include all segments *. Furthermore, in the verification phase, when a short edge element is being tried for a match, one need not search the whole model edge list



(a)



(b)

Figure 6.4.1-1 Result of minimal spanning tree algorithm (b) applied to some of the edge elements in the image of a cup and phone (a). The metric used for growing the tree is the planar distance between edge elements. See Burr and Chien (1976) for this technique.

since location of a match for the edge would only increase the CONFIDENCE by a small amount. In general, search the model more completely dependent upon the potential of the edge element to contribute to significant CONFIDENCE increase. This would introduce some error in the CONFIDENCE, but it would be traded for greater search efficiency.

A strong heuristic to implement is that at least two edge elements are present in the scene which correspond completely to two model edge elements *. This allows enforcement of edge length agreement, which reduces search in the proposer. However, this restriction was relaxed in the verifier so that broken edges can still match.

In searching a particular model for correspondence, one can take advantage of a feature called the maximum extent of the object. This is the maximum length diameter that can exist in the object. When used in searching for the second pair element in the proposer, line elements can be rejected if they lie at a position exceeding this distance. Greatest search reduction can be gotten with this heuristic if the particular test object or subobject is small in all aspects. This provides further justification for hierarchical decomposition -- to make subelements compact in shape.

Another heuristic is to limit centers of gravity of model and scene to be within a preset tolerance after the initial proposal of two edges *. This eliminates obviously

bad proposals, where the model is oriented to lie outside the scene. It is obviously more powerful the fewer the number of objects in the scene. Further, if the table top plane is known, refuse any proposals which cause an object to lie in part below the table. After one body is found in a scene, prohibit others from intersecting those already found.

7. NARROW ANGLE STEREO

For various reasons it may be desirable to represent structures with curved segments. One particular scheme would allow piecewise linear and piecewise circular primitives for edges of objects. Since central projection does not preserve curvature, as it does linearity, it is desirable to perform fitting of curved segments in space rather than in the image, hence the need for efficient computation of depth maps.

Because of the conflicting requirements of feature comparison and triangulation, a particular method normally must accept a tradeoff between the two. There has been little attention given to narrow angle approaches because of inherently poor triangulation accuracy. However, the possibility of rapid, reliable and simple techniques for feature comparison make the narrow angle technique attractive for computing depth maps. The approach taken here is to use dynamic smoothing to reduce the deleterious effects of image noise on ray intersection.

7.1 Symbolic Correlation of Edge Elements.

In this approach edges are detected and followed in each of a pair of images. Comparison of features is done with a threshold function whose arguments are derived from the edge list data. Since edge points on the list are

arranged in the order in which they were tracked in the picture, continuity can be enforced quite easily. After an initial match is found between two edge elements, the match for the next element on the list is restricted to lie in a small band (usually ± 3 units) about the last matching edge datum. This is better than in the previous tracking method (Section 5.4) where search is restricted to areas in the original picture, since edge list data (position, orientation, step size) can be easily smoothed prior to matching. This approach generally results in longer continuous 3-D contours than in method 1.

Picture pairs are vertically shifted, since camera noise components in this direction were observed to be small. The decision polynomial contains five terms:

$$\mathrm{DIFF} = C1\,|\Delta x| + C2\,|\Delta y - d| + C3\,|\Delta Dx| + C4\,|\Delta Dy| + C5\,|\Delta m|, \quad (3)$$

where (x,y) are coordinate values, (Dx,Dy) are gradient angle components (Figure A-1), m is the gradient magnitude, and d is stereo disparity. Deltas indicate differences in these quantities between the two pictures, and C1-C5 weight the effects of the feature differences. In general, if a feature is noisy or is a poor discriminator for predicting matches, then its coefficient is lower. Δx values are limited to -1, 0, and +1 because of projection constraints. Initially C2 is zero, but after a successful match, it is weighted to favor a subsequent edge match at

the previous disparity value, d. This is the 3-D continuity constraint. The 2-D constraint was imposed by restricting the search field to neighboring edges on the edge lists. A particular candidate must produce a local minimum over this field and be less than a threshold to qualify as the match. If none are found, then exhaustive search of the edge lists is successively invoked to obtain new match pairs. The results of the matching are displayed after triangulation. A rotated structure is portrayed to show triangulation and matching accuracy. It should resemble the shape of the actual object if successful.

7.2 Nonlinear or Dynamic Contour Smoothing.

To permit triangulation to be used practically for narrow angle stereo, a new method was developed for smoothing digital contours which minimizes rounding of corners and maximizes smoothing of low curvature sections of the contour. In order to adequately reduce noise, some smoothing must be allowed even at sharp corners. However, if the edges have been tracked in a similar manner in both pictures, this smoothing will be in the same direction, and thus contribute little error in the triangulation, which is inherently difference sensitive.

An approach found successful in reducing contour noise is one which attempts to find maximal length smoothing intervals at points along a contour. It also assumes that

noise is amplitude limited. At a particular edge point, intervals of equal plus and minus extent are sought such that all points within the interval are within a fixed perpendicular distance from the line connecting the interval end points. The maximum length interval which satisfies this requirement is defined as the smoothing interval for the point. At locations near contour ends, this interval must necessarily be limited so as not to extend beyond the contour. The smoothing function within the interval is just the average of all positions (also angles, or intensities) within the interval. Various weightings could be tried, but the unweighted one was satisfactory and also allows efficient implementation.

7.2.1 The WORM Smoother.

An efficient version of this algorithm takes advantage of the fact that neighboring points on a curve will have similar smoothing intervals. In addition, a running average of data can be kept and incremented, rather than recomputing averages at each new point. It is called WORM since the plotted smoothing interval when observed dynamically resembles a straight rigid worm crawling through a tunnel whose radius is the error threshold.

The implementation begins by smoothing the point next to one end using an interval of three (see Figure 7.2.1-1). A tail is defined as the pixel at the contour end, and a

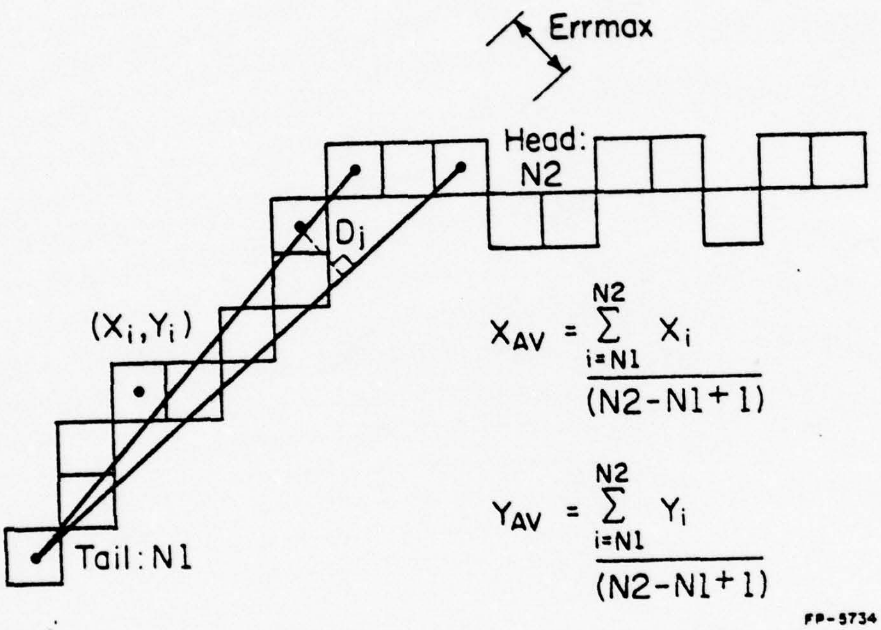


Figure 7.2.1-1 Illustration of dynamic interval determination for the WORM smoothing algorithm. The head location is advanced until there exists some D_j along the head to tail interval which exceeds ERRMAX. Then the tail is advanced until D_j is brought below ERRMAX. This process is iterated, the current interval being used to smooth the contour point at its center.

head as the opposite end. The head is advanced in position by two units and a test is made to see if any point in the interval exceeds the distance threshold from the WORM. If not, the edge point midway on this interval is smoothed relative to the given interval. The head position is successively incremented by two until such a deviation is found. Call it Dj. An increment of two always guarantees that a contour point exists midway along the interval. The tail position is now incremented in units of two (same reason) until the distance from Dj to the head-tail line falls below ERRMAX. In the process of shrinking, the center point of each interval is smoothed with the current interval as the range. The sums of x and y coordinates over the initial interval (length 3) is computed. They are incremented by adding x, y coordinates from each added point and decremented by subtracting x, y coordinates from each removed point. A running average is thus computed by dividing these values by the number of points in the interval. The process ends when the next to last point in the contour has been smoothed.

7.2.2 An Iterative Variation for High Frequency Noise.

A variant of the previous method of smoothing is based on an assumption that the noise is relatively high in frequency. If this is true, then smoothing can be accomplished by using a fixed interval and weighting, say, nearest neighbors in inverse proportion to the absolute

curvature estimate at that point. The progression of weighting values should be smooth between the extremes. A particular function satisfying these requirements is

$$WT = 1 - \cos(\alpha/2), \tag{4}$$

where $\alpha = \cos^{-1}(\vec{v1}\ \vec{v2} / \vec{v1}\ \vec{v2})$, and $\vec{v1}$ and $\vec{v2}$ are vectors defined between a contour point and those points respectively $\pm$ n removed from it. α can be interpreted as a bending angle. The smoothing is done on multiple passes over the contour, each using decreasing values of the interval n. Each pass smooths only points having the property

$$|\alpha| \geq \pi(1 - 1/n), \tag{5}$$

where n and α are defined above. Subsequent passes extend the smoothing to points closer to sharp corners, resmoothing the previous ones. The effect is ultimately to impose a low pass filter on all points, the upper cutoff of which is lower for low-frequency portions of the curve, or the flattened portions. The purpose of implementing this with several passes is to allow better estimation of the bending angle, as its value is perturbed by noise in the contour.

This last technique was the first one implemented, and was intended to be used to smooth scene contours. However, the assumption of high frequency noise was not valid for the

particular images, and a resultant cyclic noise pattern remained. The assumption of fixed amplitude of contour noise was a better one, and the WORM technique thus proved successful. Its tendency to round edges slightly more than the iterative method was not objectionable for stereo matching so long as rounding occurred similarly in each picture.

7.3 Cleanup of the Depth Map

7.3.1 Reliability Estimation.

If we look ahead for a moment and observe the results of narrow angle triangulation (Figure 9.4-5a), we see that long straight edges get processed reasonably well, whereas curves and corners have greater error. Furthermore, due to the problems associated with edge orientation in a single pair of images, poor registration results when edge orientation is parallel to the image shift.

This can be corrected by smoothing, but can best be done if an estimate is made of such error. This estimate can then be used to guide a dynamic smoothing algorithm. Unless long straight edges in the image are aligned in the same direction as the image shift between stereo pairs, then large errors as discussed will occur only over relatively short intervals. It is usually possible to orient the views so that this condition is met except for extreme cases.

Since smoothing error varies directly with contour curvature, and orientation error with |edge slope| relative to the horizontal, then we attempt to estimate depth error as follows (geometric distortion is not considered for reasons discussed earlier):

$$\text{ERR} = K*ec + es, \tag{6}$$

where ec and es are the error estimates respectively due to curvature and slope:

$$ec = \sqrt{\begin{array}{l}(\ (\ x(j-2) + x(j+2)\ )\ /\ 2 - x(j)\ )^2 \\ +\ (\ (\ y(j-2) + y(j+2)\ )\ /\ 2 - y(j)\ )^2\end{array}} \tag{7}$$

$$es = |(\ y(j+2) - y(j-2)\ )\ /\ (\ x(j+2) - x(j-2)\ )| \tag{8}$$

x(j) and y(j) are the image coordinates indexed along the contour. The coefficient K is required to scale the relative effects of ec and es. A value of 25 was found adequate for depth smoothing.

7.3.2 Hysteresis Smoothing.

There may be several attractive ways to use this information to optimally smooth the disparities, but one successful way consists of a technique known as hysteresis

smoothing. In this method, as normally implemented, elements along a curve remain unchanged except when the difference between one point and the next exceeds a threshold. In this case the element is changed to the value of the last visited one. This lag effect is continued until once again consecutive points are within the specified tolerance. Since an external estimate of error is available (Equation 5), truncation can be invoken when this estimate exceeds a threshold. Since error becomes appreciable only over relatively small intervals, the truncated value is likely better than the original estimate. Disparity is smoothed rather than 3-D position, since the picture plane components of the 3-D coordinates are usually better defined than the ray coordinate, or disparity. Only in orthogonal projections are the picture plane and depth components independent, so in general it is better to smooth the disparities before triangulation. An alternate to truncation would be linear interpolation within the error intervals.

7.3.3 WORM Smoothing.

Figure 10.4-6 shows the result of hysteresis smoothing of Figure 10.4-6a. The original 3-D data is considerably improved by this simple process. Depth errors have been reduced so that all disparities have roughly fixed maximum deviation. This is precisely the condition required for WORM smoothing, so a good suggestion would be to use it for

further improvement. Since the error is no longer strictly of fixed maximum deviation, but only nearly so, the WORM smoother is modified to account for this. The modification estimates smoothing intervals based on least squares fitting of successive increments of the contour, and error is measured as rms deviation of contained points along the interval. The head and tail movement are now controlled according to whether the error is greater or less than a threshold. Since there is no longer a guarantee that tail shortening will terminate, the restriction that it stop when the head-tail interval reaches 3 is included. In this way, if certain portions of the curve have individual local rms errors exceeding the preset threshold, the smoothing interval will be set at the minimum of 3 and advanced appropriately until normal head-tail movement can be resumed.

8. CIRCULAR ARC APPROXIMATIONS.

8.1 Extension of Ramer's Method.

An extension of the iterative endpoint fit (Ramer (1972)) was made adapting it to circular arcs. Instead of connecting two endpoints of a contour, an additional point roughly midway between the two endpoints is added, and the unique circle passing through these three points is found. Edge point errors are measured radially to the estimated arc and the maximum error determines a new segmentation point for recursive entry, similar to the linear method. Arc estimates are based on single points, and thus a single wild point can cause a splitting. Its scope is thus limited to contours with relatively low noise. Prior smoothing with one of the filters described, though, should help considerably.

8.2 Centroid Method.

A better method, which does not make such strong requirements on curve noise is a heuristic one which takes advantage of roughly uniform spacing of edge points along a contour. A fortunate circumstance is that both nonlinear smoothers described do just that -- arrange points with locally uniform spacing. However, even when this is not valid, the method provides acceptable curve fitting, though,

yielding several smaller arcs where a single one might have sufficed.

First the technique for fitting a single arc to an entire chain is described. The contour (Figure 8.2-1) is divided into three consecutive segments, each having equal number of points (or nearly so). The centroid of each third is computed and the unique circle intersecting these three points is found. Though this circle generally will not be a good fit for the set of points, its center is a good estimate for the center of the best fitting arc. The best fit radius is then defined as the average distance from all contour points to that center point. The resultant arc thus has equal weighting of points inside and outside. An error can be computed as a function of the radial distances from the contour points to the fitted curve. The arc endpoints are defined as the intersection of the fitted arc with the lines from each endpoint of the contour to the arc center. The method is fast and generally yields good fits. Both it and the iterative endpoint method are applicable to either 2-D or 3-D contours. In 3-D the three centroids also define the plane in which the arc lies. It may be preferred over other techniques, since the arc is not restricted to intersect contour points.

This method is easily adapted to fitting of multiple arcs along a contour. Initially several points at one end of the contour are fitted using the method. If the computed

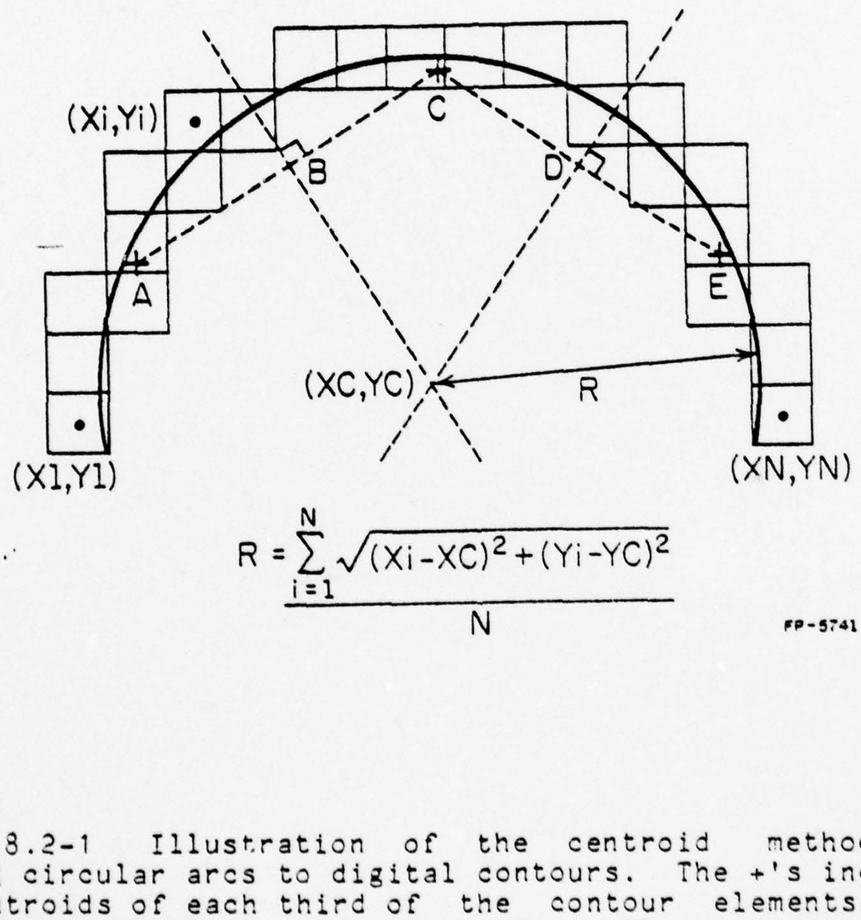


Figure 8.2-1 Illustration of the centroid method of fitting circular arcs to digital contours. The +'s indicate the centroids of each third of the contour elements, and (XC,YC) is the point equidistant from all three centroids. The radius of the approximated arc is the unweighted average of contour point distances to (XC,YC). For 3-D curves the centroids also define a plane which contains the fitted arc.

error exceeds a threshold (preassigned based on goodness of fit requirement) then the last fitted arc is taken as the best fit over its interval. Otherwise the interval is extended by adding points to it. If at any time, the error of the starting interval exceeds the minimum error, then its arc fit is taken as final over that interval. However, this usually indicates a poor initial tolerance estimate.

9. RESULTS

9.1 Equipment.

Images were obtained from a Philips Norelco low-blooming silicon vidicon. An entire picture is converted to 9996 PDP-10 words in one-half of a television frame period, or 1/60 second. Pictures contain six bits of grey scale, at each of 252 by 238 resolution units. For research purposes (comparing results of different algorithms) all data shown was obtained from pictures stored on disk. However, the system can be used in real time. Jones (1975) used this feature to track moving objects, employing the edge detection operator described here.

A turntable was built for obtaining the various views needed for stereo processing (see Figure 9.1-1). The table has two axes of rotation, a swivel axis (vertical) and a tilt axis (horizontal). Various positions of the table were adjusted manually and viewing angles measured by hand. Although a system utilizing these ideas might consist of a multiple-camera or a multiple-mirror configuration, it was felt that no loss of generality would be had if the turntable system were used for experiments. In addition it would allow flexibility in configurations of views that others might not permit.

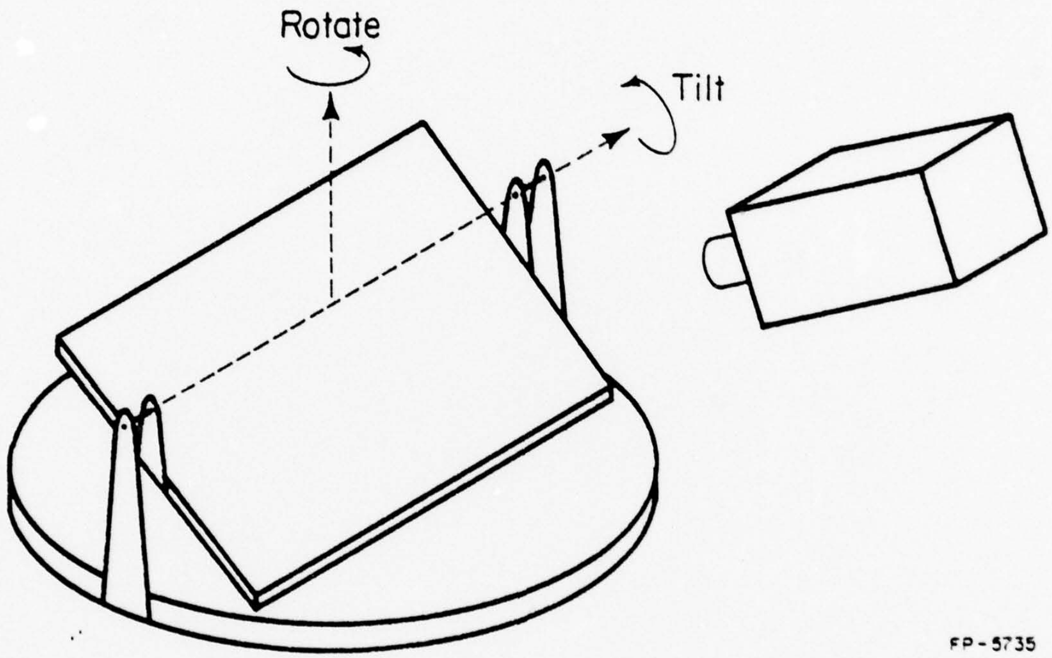


Figure 9.1-1 Illustration of the turntable used for experiments. Rotation and tilt axes are orthogonal and coincident at one point.

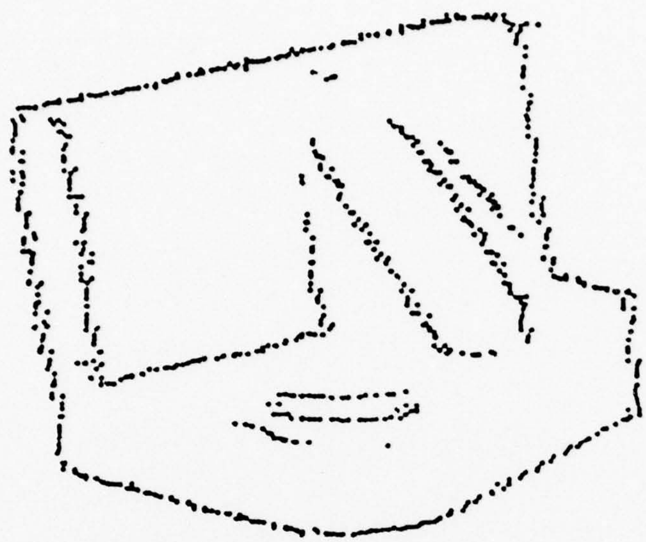
The first tests were done on simple objects consisting of curved and straight edges, and fabricated from balsa wood. Subsequent tests were done on common objects found about the lab.

All programs were written in BLISS-10, an ALGOL-like language for implementing system software on the PDP-10 computer.

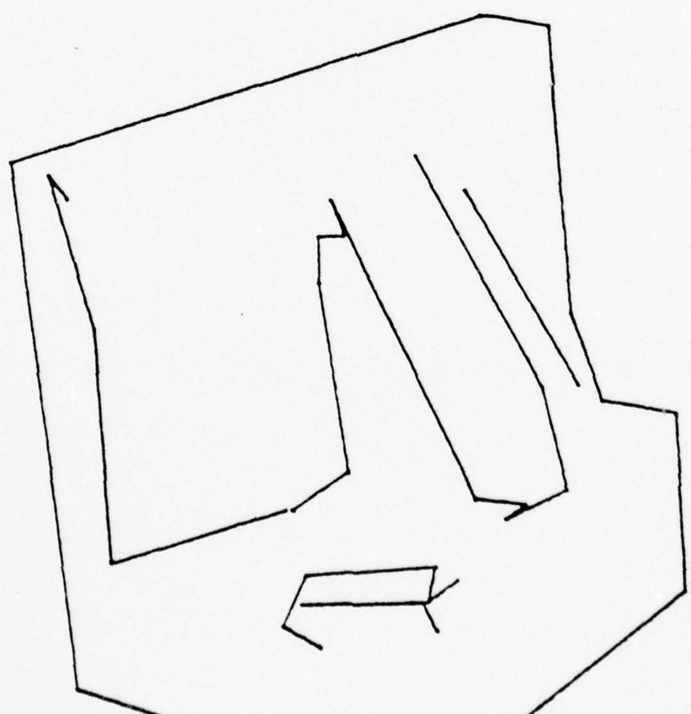
9.2 Multiple Views.

In Figure 9.2-1 we see an object with an angle brace and a hole in the left foreground. The top illustration (Figure 9.2-1a) shows the result of the edge follower operating on an image of the object. Note the greater noise on vertical edges relative to horizontal edges. The edge positions shown are those computed after parabolic interpolation in the in x and y directions, so any observed noise is primarily external or unrelated to pixel sampling.

The lower illustration (Figure 9.2-1b) shows the result of fitting these edge contours with line segments using the iterative endpoint method. An error criterion of 2.5 pixels was used in the fitting. The indicated image corresponds to the center picture of a three picture set. 3-D coordinates were computed at the endpoints of each line segment shown, and the 2-D wire frame structure projected to form a 3-D structure (Figure 9.2-3).



(a)



(b)

Figure 9.2-1 (a) Edge picture of an object. (b) Linear approximation of edges in 'a'.

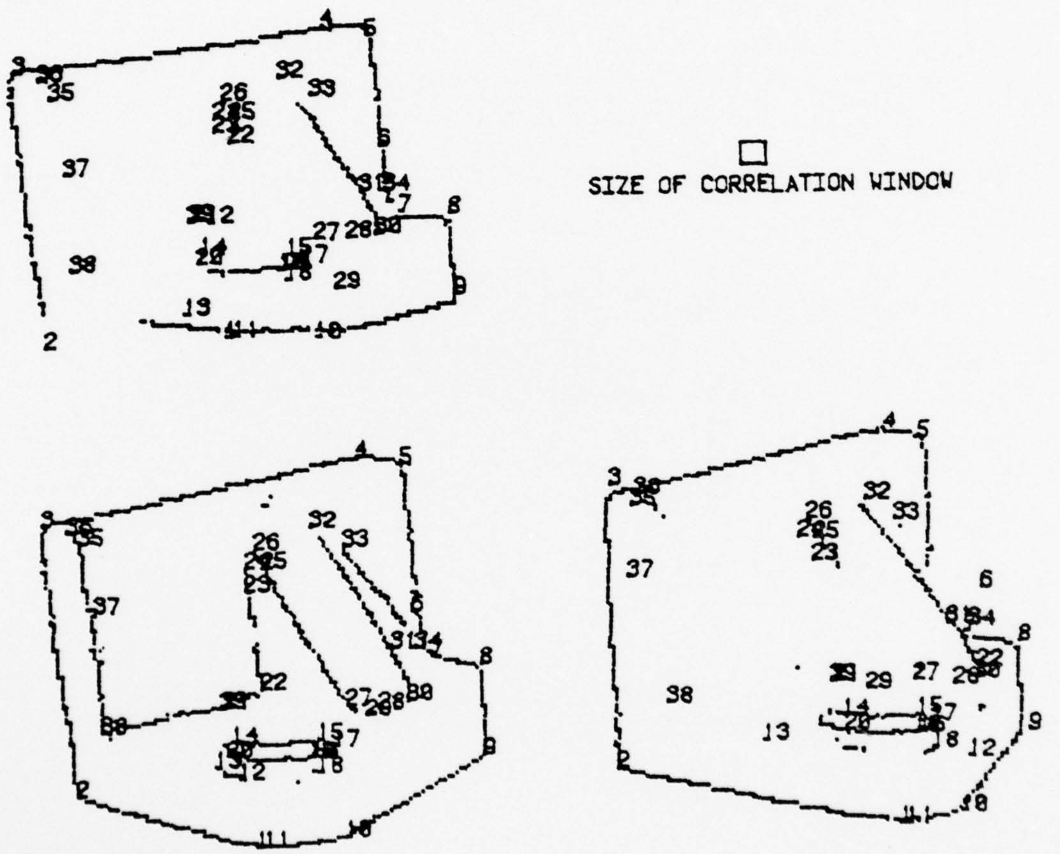


Figure 9.2-2 Results of the bulk correlation process on three views of the object in Figure 9.2-1. Numbers are arbitrarily assigned to indicate regions that were matched by the program.

Figure 9.2-2 shows the result of the comparison technique on the three pictures of the set. The lower left picture is the one shown in Figure 9.2-1. The north view is shown in the upper left corner, and the east view, in the lower right. Numbers are arbitrarily assigned to the various line ends of Figure 9.2-1b. The appearance of an identical number in the other views indicates that location was matched with the feature from the center view. Correlation was performed on the original images and not the edge pictures as shown.

Figure 9.2-3 shows four projections of the computed 3-D structure obtained by triangulating corresponding points from Figure 9.2-2. The structure has been rotated 30 degrees in each of four directions (NESW) and projected into a simulated viewing plane. The gross structure seems to indicate the proper shape of the object, though some of the smaller details (hole) are distorted.

Figure 9.2-4 shows another example for a screwdriver leaning at an angle on a small box with an open lid. Figure 9.2-4a represents the edge picture, and Figure 9.2-4b, the segmentation of 'a'. Figures 9.2-4c and 9.2-4d show the computed 3-D structure as viewed from two different angles via simulated projection. Note that it seems mostly correct except for two lines in the lower portion of the box. The reason for the poor match here is likely due to occlusion, since the lowest value of the correlation function is chosen

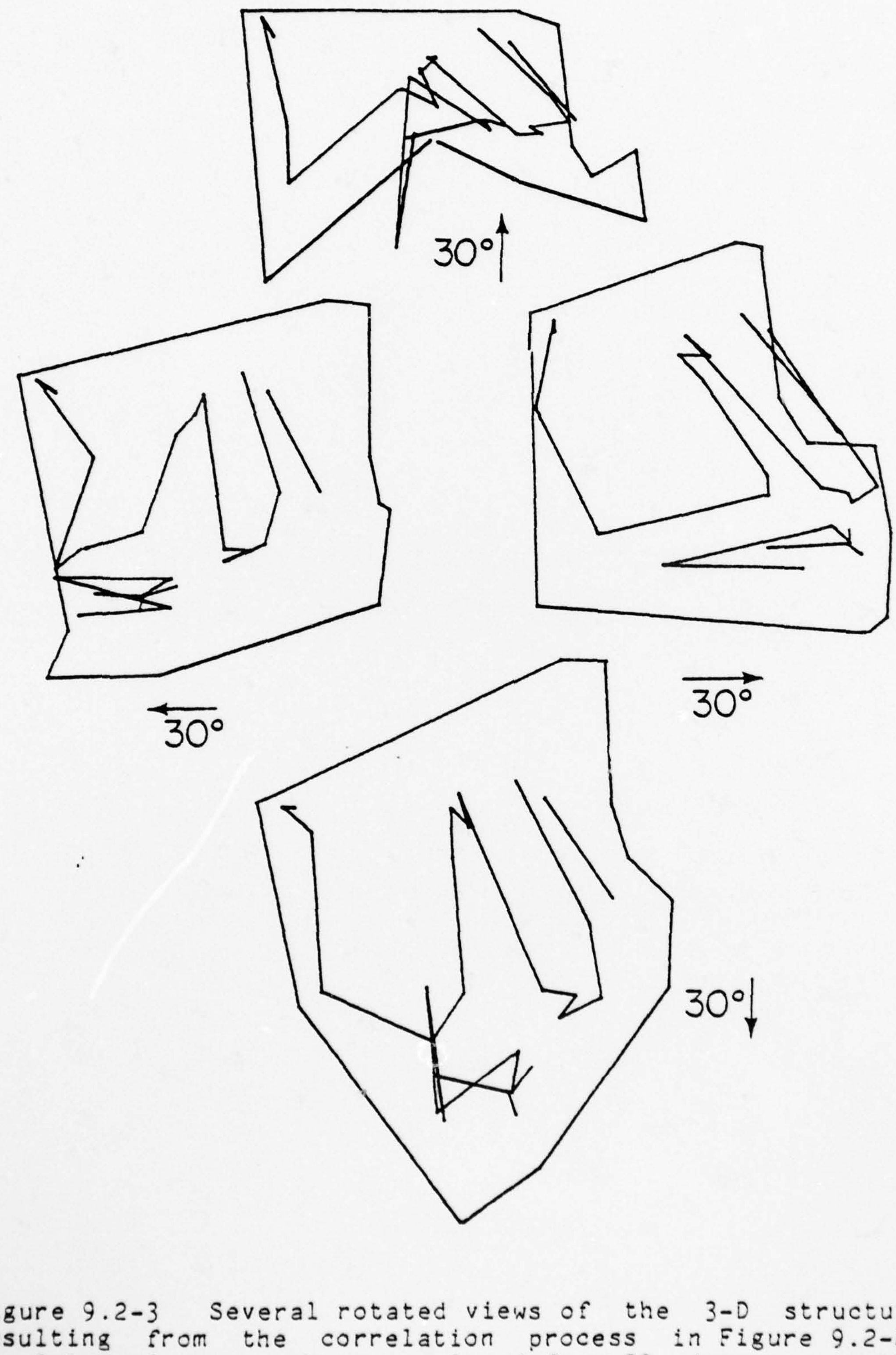
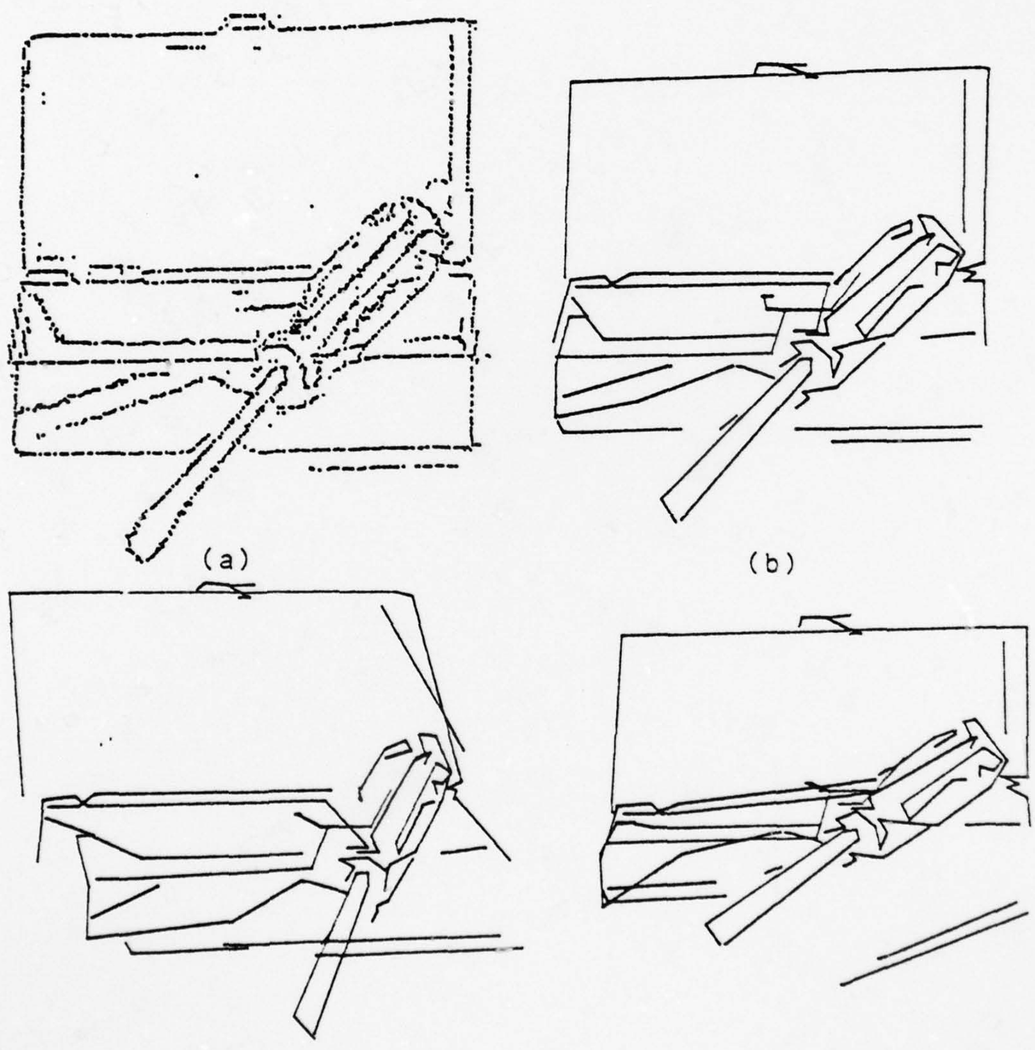


Figure 9.2-3 Several rotated views of the 3-D structure resulting from the correlation process in Figure 9.2-2. Simulated views are shown as viewed from 30 degrees north, south, east, and west of the center view (about the turntable axes).



(a) (b) (c) (d)

Figure 9.2-4 (a) Edge picture of a screwdriver and open box. (b) Linear approximation of edges in 'a'. (c) View of the structure computed from 'b' and three stereo views. (d) Another view of the same structure.

AD-A057 650 ILLINOIS UNIV AT URBANA-CHAMPAIGN COORDINATED SCIENCE LAB F/G 14/5

ON COMPUTER STEREO VISION WITH WIRE FRAME MODELS.(U)

DEC 77 D J BURR DAAB07-72-C-0259

UNCLASSIFIED R-805 NL

2 OF 2

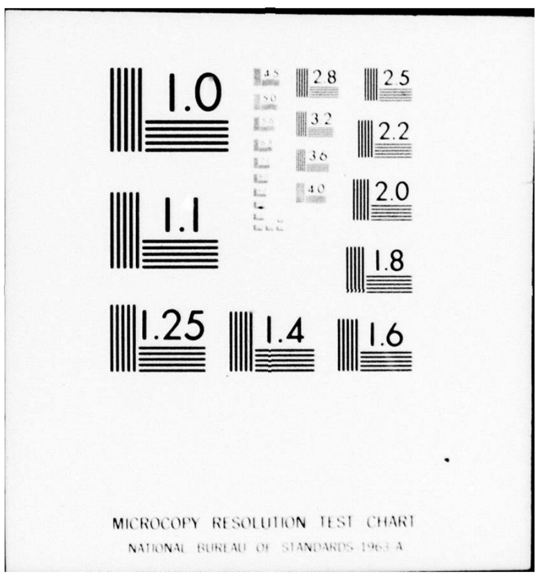
AD A057 650

END

DATE FILMED

9-78

DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

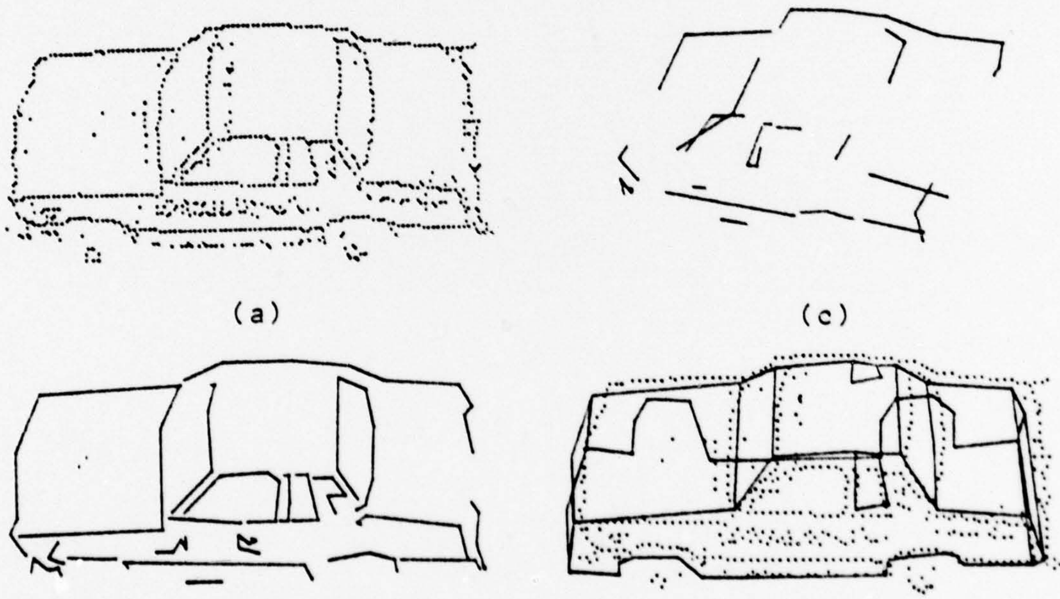
at all times, even if the corresponding feature is obscured in the other view. An improvement would be to perform two correlations at points slightly removed from a corner when a depth discontinuity is anticipated, or at all junctions, followed by a discontinuity test.

Figure 9.2-5 shows the result of computing depth incrementally along the edges of Figure 9.2-1a, using the multiple view method on the three images indicated from Figure 9.2-2. Note the aberrant segments arising from improper matching of some edge features along the boundary.

9.3 Matching of 3-D Structures.

Matching of the three-dimensional structure of a car image is shown in Figure 9.3-1. The computed structure is shown rotated in Figure 9.3-1c, after some edges highly sloped relative to the image plane were removed. This structure was computed from three car images, but without the aid of the redundancy of the multiple view correlation product. Instead, the edge direction at a line end was computed, and the east or north picture was chosen for pairwise correlation based on whether the edge direction was greater than 45 degrees or less than 45 degrees, respectively. The resulting greater frequency of match errors necessitated the use of some clean up in the 3-D structure. This technique was the first one tried, redundancy later being provided by definition of composite

Figure 9.2-5 Example showing incremental edge tracking using the multiple view process. Edges are tracked in three dimensions by searching in the vicinity of the last found edge element until the edge leaves the search interval. When this occurs, search is resumed on a global basis to relocate the lost element. Constrained search is resumed when the edge is relocated.



(a) (c)

(b) (d)

Figure 9.3-1 (a) Edge picture of a car. (b) Approximation of edges in 'a' with line segments. (c) Rotated view of 3-D structure as computed from 'b' and three stereo views. Some abberrant highly sloped depth edges have been filtered out. (d) Best match of the car model to the 3-D scene in 'c'; Confidence=0.713; PDP-10 run time=13 seconds.

correlation.

The hand encoded wire frame model of the car was tested against this structure, using the matching program. The correct orientation was found after 13 seconds of PDP-10 cpu time and is shown in Figure 9.3-1d. Note that the match was possible even though much of the detail of the 3-D car structure was absent. This demonstrates power of 3-D constraints for a complex shape with missing information, segmentation irregularities, and depth errors.

The first shapes tested with the program were simple ones fabricated in the lab. The matching results of some of these are shown in Figure 9.3-2. In each case only one or two intermediate tries were made in the process of optimizing the confidence value of the match. The recognition program was run using only the model of the object in the scene. The intermediate 3-D structures obtained for these images are not shown. Notice that objects which have predominantly curved edges are properly oriented (Figure 9.3-3). Even though the actual segmentation produced by the program is not likely the same as that produced by hand encoding the model edges, this did not seem to affect recognition. Notice also that occlusion does not prevent recognition of the occluded object (Figure 9.3-4). In this example the entire scene was matched against each of the two objects in succession, without removing any edges. This exhibits the power of the

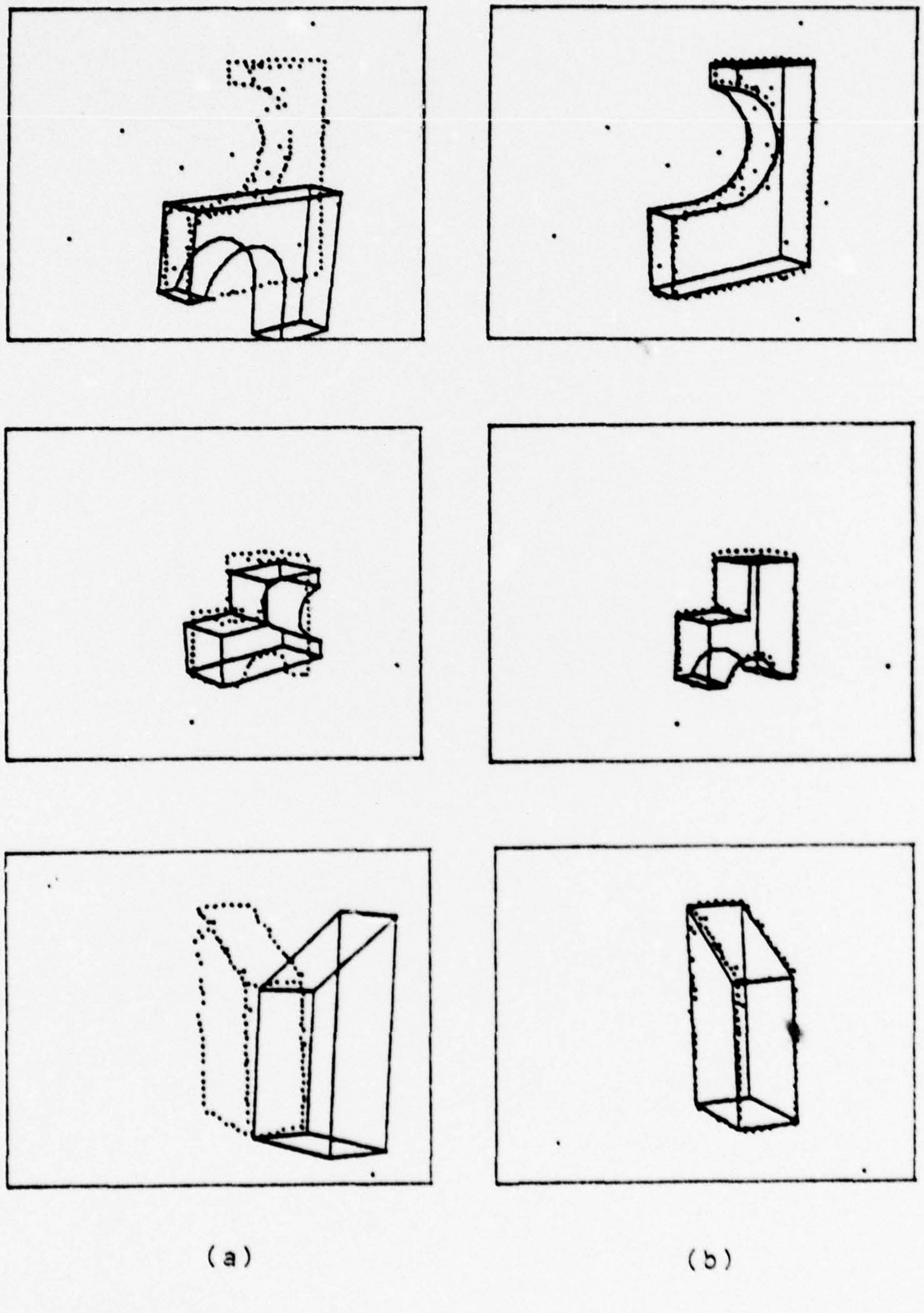
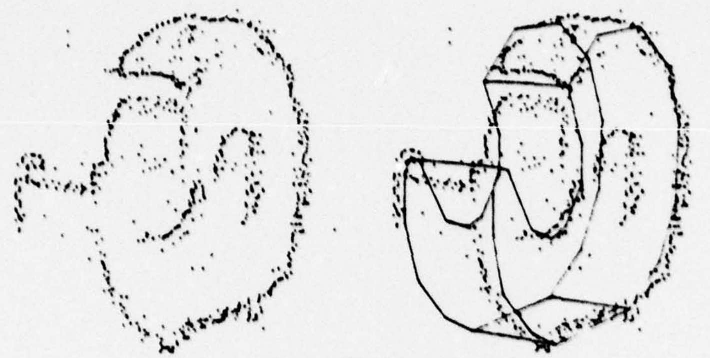
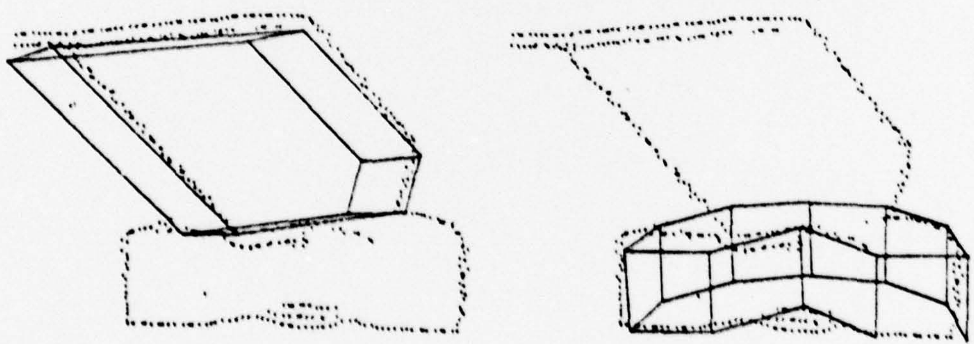


Figure 9.3-2 (a) Intermediate stages and (b) final stages in the model matching program for some simple scenes. Three dimensional scenes (not shown) obtained via the multiple view method were the input to the model matcher. Note the ability of the program to match subfeatures in the process of finding the correct orientation.



(a) (b)

Figure 9.3-3 Result (b) of matching an object (a) with curved edges (tape dispenser). Presence of noise and segmentation irregularities between image and model pose no real problem, provided there is sufficient evidence to determine a match.



(a) (b)

Figure 9.3-4 Examples of finding some simple objects in an occluded scene. (a) The wedge-shaped object was found correctly when its model was tested against the entire scene. (b) The crescent shape was found when tested against the entire scene, even though partly occluded. Efficiency could be gained by removing edges from the scene that were previously identified (wedge elements).

geometric constraints in rejecting parts of a cluttered scene inconsistent with the current model.

Since matching efficiency is directly related to the accuracy and completeness of the 3-D scene, more complex scenes were not tested. Furthermore, scenes had to be restricted to the center region of the camera to minimize distortion. It was felt more desirable at this point to focus attention on better ways to compute 3-D structures from scenes. It is felt that the ability to match such structures has been sufficiently demonstrated, and that improvements toward higher scene complexity and search efficiency should necessarily focus on the 3-D depth extraction process.

9.4 Narrow Angle Stereo.

The results shown here are based on comparison of two pictures. The unprocessed edge picture of the image is shown in Figure 9.4-1. It consists of a screwdriver in the foreground and a pair of scissors in the background. Notice in Figure 9.4-2 an enlarged section of the scissor blades which shows in detail both the positions of edge segments and their orientations as computed from the bidirectional gradient. The observed jagged nature of the positions is primarily due to external noise, and is not attributable to pixel quantization. This is an example of the fine discrimination capability of the edge operator for measuring

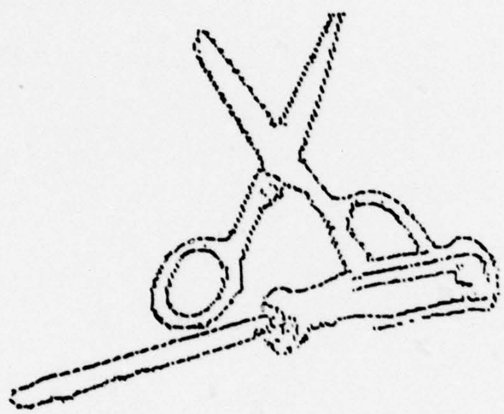


Figure 9.4-1 Unprocessed edge picture of scissors and screwdriver.

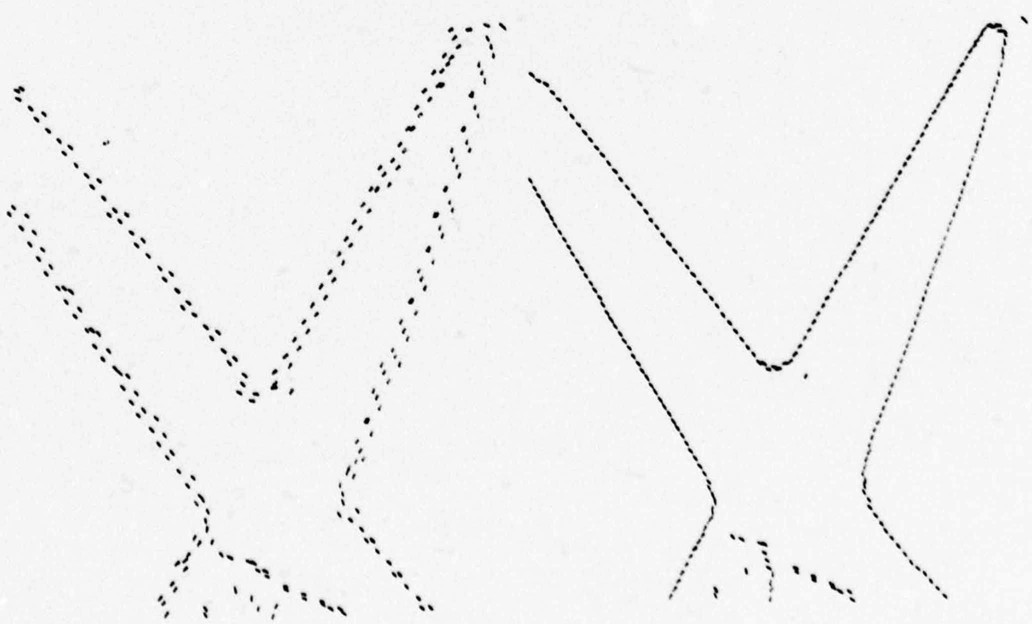


Figure 9.4-2 (a) Expanded view of scissor blades. The segment orientation indicates the local edge direction at that location. (b) Same view after processing with the WORM smoother.

edge detail. The edge directions are not so noisy as the positions, since they are determined from x and y differences which have been averaged over several pixels. The contour follower would likely give excellent results on pictures with less noise.

Compare Figures 9.4-2, a and b, and notice the nearly complete elimination of all edge position noise. The presence of a slight high frequency variation is due to the fact that the dynamic smoothing was done only on odd length intervals. This high frequency residual was subsequently removed with a simple linear smoother which averaged points with 1/2 the values of each nearest neighbor.

Figure 9.4-3 shows an overlay of the two edge pictures used in the stereo matching after being processed with the WORM smoother. Notice the fine detail in the disparities between corresponding features, especially in the screwdriver and scissors blades. In Figure 9.4-4 an enlarged section near the blades is seen.

It is difficult to display the result of only the matching portion of the narrow angle method on these edge pictures. Therefore, the entire result is shown after matching and triangulation. Figure 9.4-5 shows simulated projected views of the computed 3-D structure. Figure 9.4-5a contains the raw data, 'b' the data after hysteresis smoothing, and 'c' after processing with the WORM smoother (Section 7.3.3).

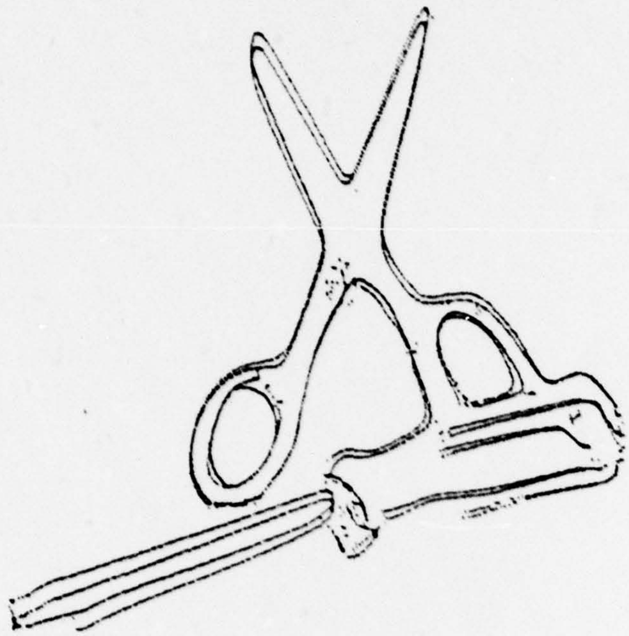


Figure 9.4-3 Superposition of two narrow angle edge pictures after edge smoothing.

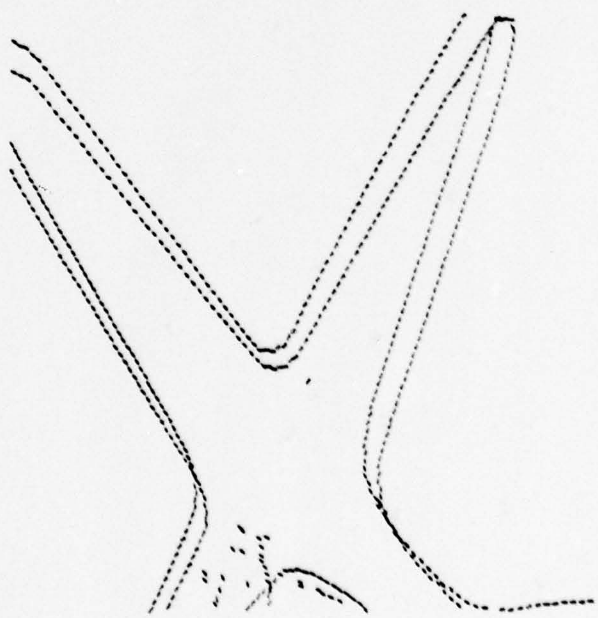
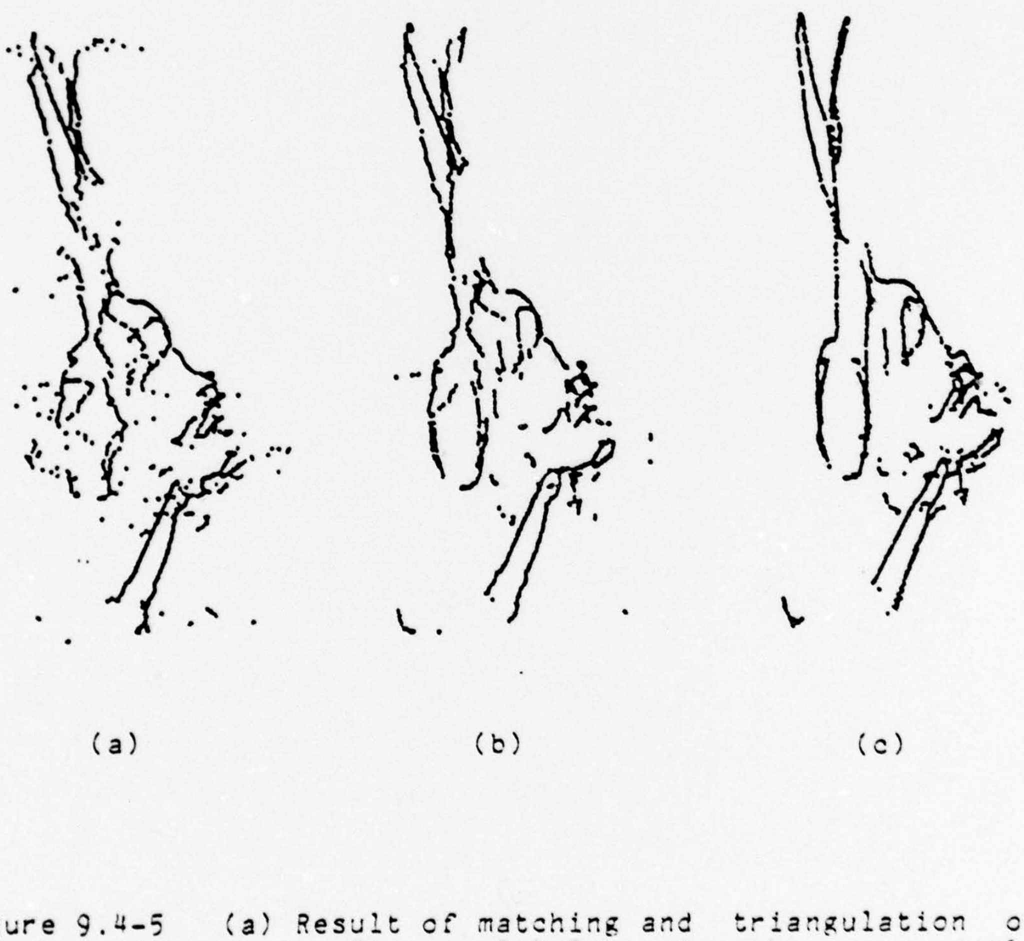


Figure 9.4-4 Superposition of the two narrow angle views of the scissors blades shown expanded after smoothing.



(a) (b) (c)

Figure 9.4-5 (a) Result of matching and triangulation of the two images in Figure 9.4-3, using the narrow angle method. The simulated view of the 3-D scene is approximately 72 degrees about the turntable center, to the left of the original views. (b) Hysteresis smoothing of the 3-D scene in 'a'. (c) WORM smoothing applied to the 3-D scene in 'b'.

The results of the high frequency dynamic smoother are shown in Figure 9.4-6. In Figure 9.4-7 observe the effects of interval propagation with multiple iterations. Figure 9.4-8 shows this smoothing technique applied to the machine part. Because it left a residual low frequency ripple, the method was discarded for smoothing the images in question. However, it should prove to be attractive if the only noise present is quantization or pixel noise, since it tends to be predominantly high in frequency (Bennett and MacDonald (1975)).

9.5 Contour Approximation.

In these examples (Figure 9.5-1) contours were fitted iteratively with circular arcs using the centroid method. The error was computed as the root mean fourth power of deviations from the arc, measured radially. The first example shows a low noise curve with fitting tolerance set to 0.5. The second shows a contour with larger noise. The fitting tolerance was 800.0. The technique was designed explicitly for fitting 3-D contours obtained with the narrow angle matching program, but can be used for 2-D fitting of picture edges.

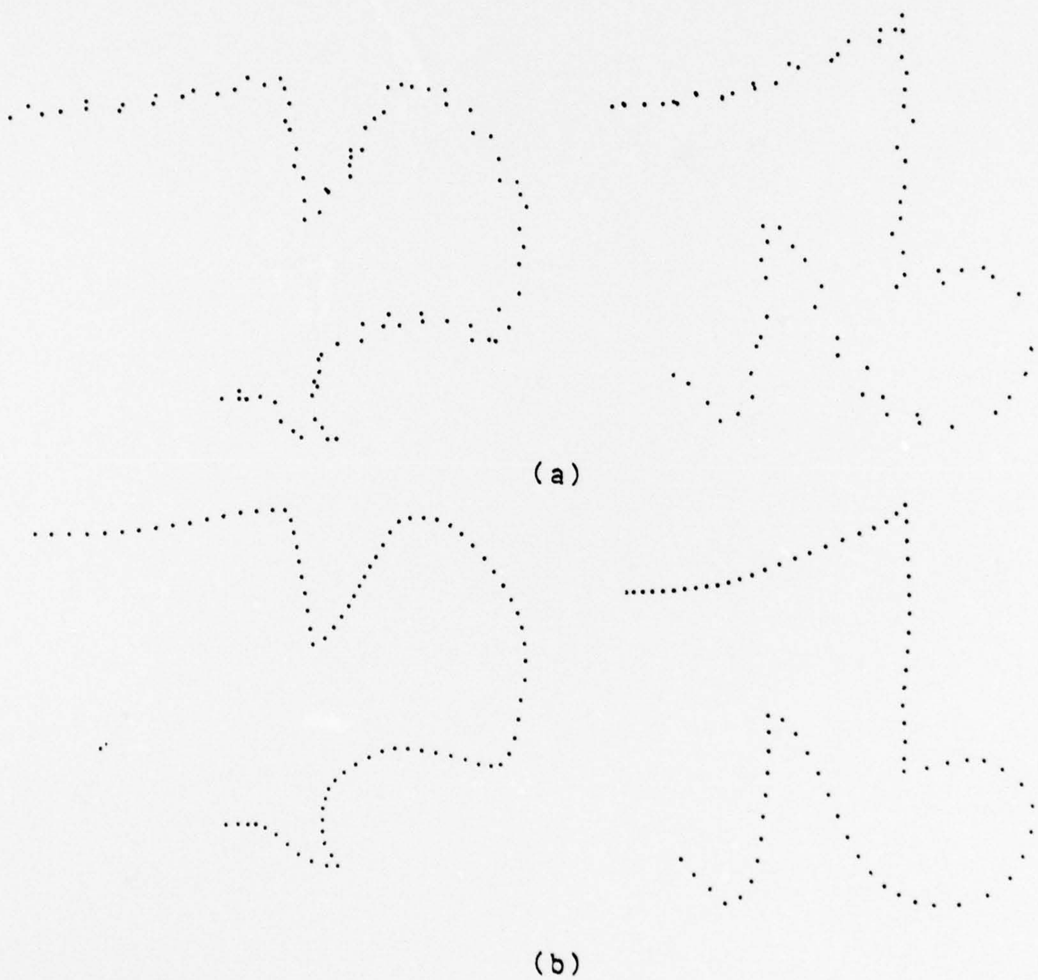


Figure 9.4-6 (a) Two contours containing some high frequency noise. (b) Contours in 'a' after being processed with the high frequency dynamic smoother. Notice the retention of sharpness at the corners. This cannot be achieved so well with linear smoothing.

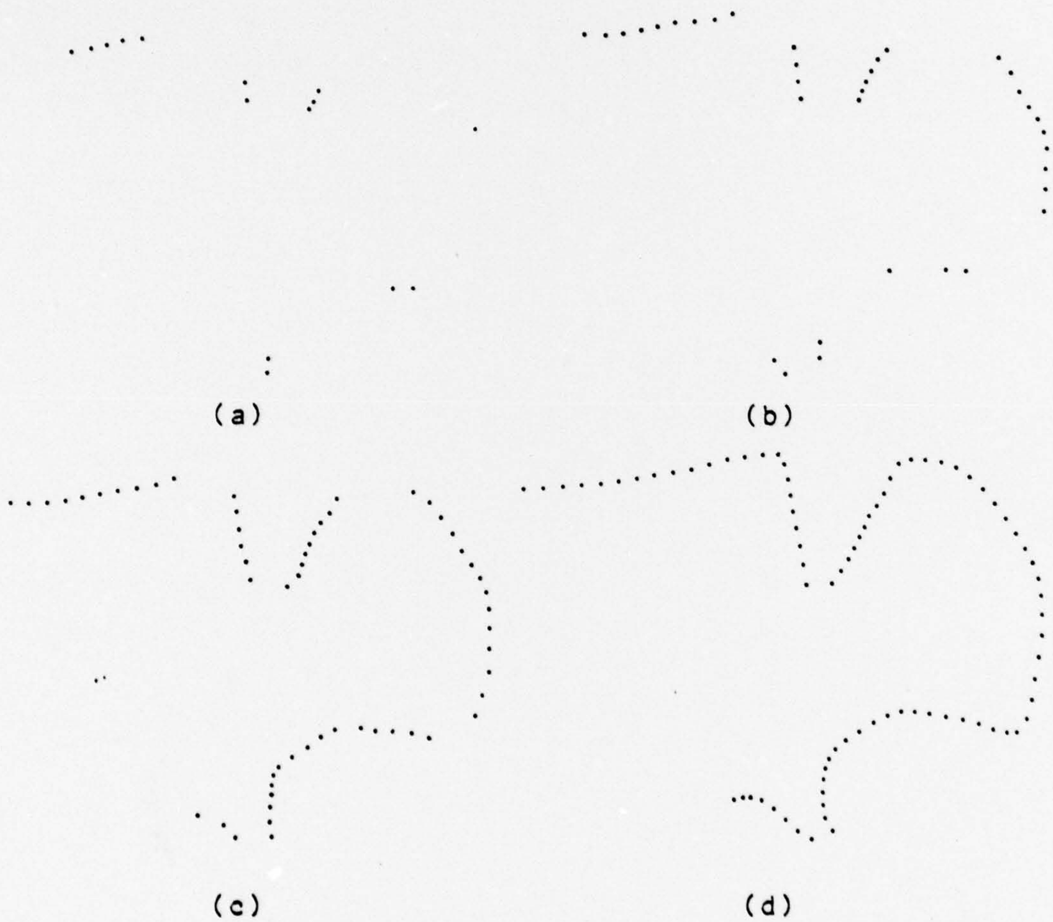


Figure 9.4-7 Example showing successive iterations ('a' through 'd') of the high frequency dynamic smoother.

Figure 9.4-8 Result of processing Figure 9.2-1a with the high frequency dynamic smoother.

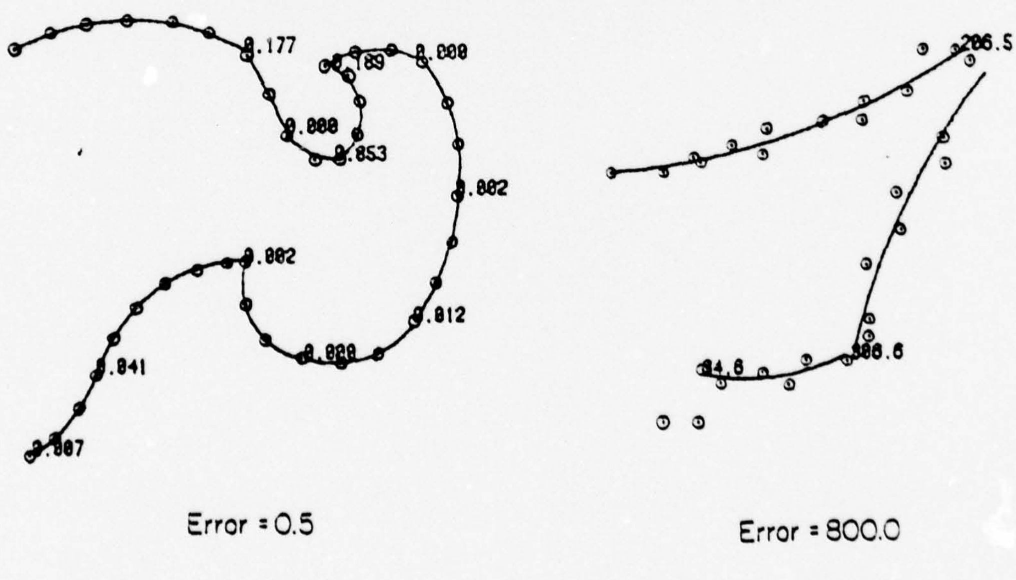


Figure 9.5-1 Output of the arc fitting routine (centroid method) for two contours with different noise amplitudes. The error measure is the mean fourth power of radial deviations from the arc. Error values are shown for each fitted arc.

10. EXTENSIONS

An extension would be in the direction of improved accuracy of depth features, say by correcting for geometric image distortion. Since it was possible to demonstrate model matching, correction of distortion was not investigated. Improvements in accuracy would result in better discriminating capability and faster, more efficient matching of objects. Furthermore, with the advent of the new CCD imagers, little or no geometric distortion is observable, thus making these techniques readily attractive.

Improvements in the direction of less wordy representations of objects have already been begun as well as strategies for 3-D segmentation with circular arc primitives. These extensions would necessarily radiate from an efficient means for computing edge depth maps, such as the narrow angle technique described. Strategies for comparing circular arcs would also need to be developed.

As Turner has suggested, hierarchical decomposition of object models is a strategy useful in implementing efficient search of a model data base, and some experimental verification of this would be desirable in the realm of 3-D prototype matching. In a similar manner, systematic means for scene decomposition to localize matching in images and model data bases (e.g. by relaxation labeling) would be desirable along with experimental verification.

Vertex-based techniques applied to minimal spanning tree segmented images were also suggested, but work needs to be done to verify the validity of this approach on body finding in real scenes. For a class of objects, at least, on a smooth background, a variant of this approach based on boundary vertices and region primitives has been demonstrated (Burr and Chien (1976)).

Extensions toward use of local color and texture features in improving model matching efficiency would be interesting. This might be all that is needed to make the shape matching fast enough for a practical vision system. By using color information simple hill climbing techniques might be quite powerful for recognizing certain objects. Furthermore, generalization of hill climbing to allow dynamic error functions would be desirable. This might result from studies relating the discriminability of object features to the estimation of translations and rotations, and would also be useful in extending iterative techniques for stereo comparison. Care should be exercised in the use of feedback of any kind (also in RL), since instability or oscillation can result.

In addition, greater use of connectivity information would be desirable in increasing shape matching efficiency. However, better reliability in depth structures might make connectivity easier to enforce. Such improved structures would prompt further research in the area of automatic

learning of shape descriptions.

11. CONCLUSION

A consistent approach for implementing multiple views in bulk correlation processes has been proposed and tested. The method has application to scenes consisting of objects with smooth surfaces, or predominantly man-made objects. The method is attractive in that matching reliability can be increased by merely adding another camera or view. It compares favorably with other approaches to matching of objects with smooth features and it works on real images. Primary advantages are its speed, and its independence of global feature requirements (segmentation irregularities) at the stereo matching level. The method is also attractive since hardware costs are continually falling, whereas the high level feature matching problem for complex scenes is yet unsolved.

A solution has been presented for representation of objects with curved edges, and for matching of such structures to three-dimensional models based on geometric constraints. It has been successful in finding 3-D locations and orientations of objects in visual scenes, even in the presence of occlusion, missing and extraneous information, and errors in stereo matching and triangulation. Its success is due to the exploitation of object-specific geometric features to disambiguate local uncertainties in stereo correlation and image feature

extraction. Techniques have been proposed for increasing matching efficiency in occluded scenes with many models. Coupled with such techniques and/or improved image sensors, the multiple view and model matching processes serve as a robust basis for object recognition in practical scenes.

Extensions to this work have been begun in the direction of improved 3-D feature determination and model representation with circular arc primitives. It is based on development of efficient schemes to construct incremental depth maps of scene edges so that curve fitting can be done in three dimensions. A fast and efficient method has been proposed and tested for comparing edge chain features using a narrow angle of view. In conjunction with this approach a dynamic smoothing technique was developed to remove much of the noise from edge chains so that triangulation can be done accurately at narrow viewing angles (2-3 degrees and less). It compares favorably with the tracking approach to narrow angle stereo (Nevatia (1976)), since many images are not required. Since edge extraction and smoothing can be performed in hardware, this was felt to be a wise tradeoff. The technique has been successful and is expected to be useful for piecewise-curved object description and matching. A start toward this goal has been achieved in the implementation of an efficient technique for fitting circular arcs to 2-D and 3-D contours. An additional method for arc fitting has also been proposed, as an an extension of a currently popular method for recursive fitting with

linear segments.

In general, the power of the two stereo techniques is attributable to the successful integration of local correlation measures with global shape information. In the multiple view technique this is demonstrated through object modeling, and in the narrow angle technique, by contour smoothing and continuity implementation. This natural interaction of low and high level processes is a desirable feature in general for cognitive systems dealing with imperfect data.

The work has been successful in many respects. It is hoped that these findings will promote some interest in stereo computer vision as a solution to tedious inspection and monitoring problems, and as a technique useful in robotic systems.

REFERENCES

Agin G. J., "Representation and Description of Curved Objects," Stanford Artificial Intelligence Laboratory Memo AIM-173, Stanford University, October, 1972.

Attneave F., "Multistability in Perception," Scientific American, Vol. 225, 62-71, 1971.

Baker H., "Building Models of Three-Dimensional Objects," (Thesis Draft) University of Edinburgh, April, 1975.

Baker H., "Three-Dimensional Modelling," Proceedings of the Fifth International Joint Conference on Artificial Intelligence, 649-655, M. I. T., Cambridge, Mass., August, 1977.

Barnea D. I. and H. F. Silverman, "A Class of Algorithms for Fast Digital Image Registration," I. E. E. E. Trans. on Computers, Vol. C-21, No. 2, 179-186, February, 1972.

Barrow H. G., Tenenbaum J. M., Bolles R. C. and H. C. Wolf, "Parametric Correspondence and Chamfer Matching: Two New Techniques for Image Processing," Proceedings of the Fifth International Joint Conference on Artificial Intelligence, 659-663, M. I. T., Cambridge, Mass., August, 1977.

Baumgart B., "Geometric Modeling for Computer Vision," Stanford Artificial Intelligence Laboratory Memo AIM-249, Stanford University, October, 1974.

Bennett J. R. and J. S. MacDonald, "On the Measurement of Curvature in a Quantized Environment," I. E. E. E. Trans. on Computers, Vol. C-24, No. 8, 803-820, August, 1975.

Burr D. J. and R. T. Chien, "The Minimal Spanning Tree in Visual Data Segmentation," Proceedings of the Third International Joint Conference on Pattern Recognition, 519-523, Coronado, Calif., November, 1976.

Burr D. J. and R. T. Chien, "A System for Stereo Computer Vision with Geometric Models," Proceedings of the Fifth International Joint Conference on Artificial Intelligence, M. I. T., Cambridge, Mass., August, 1977.

Chang Y. H., "Machine Perception of Objects with Curved Surfaces," Coordinated Science Laboratory Report R-641, March, 1974.

Davis L. S. and A. Rosenfeld, "Applications of Relaxation Labeling 2: Spring-Loaded Template Matching," Proceedings of the Third International Joint Conference on Pattern

Recognition, Coronado, Calif., November, 1976.

Della Vigna P. and F. Luccio, "Some Aspects of the Recognition of Convex Polyhedra from Two Plane Projections," Information Sciences 2, 159-178, 1970.

Duda R. O. and D. Nitzan, "Low Level Processing of Registered Intensity and Range Data," Proceeedings of the Third International Joint Conference on Pattern Recognition, 598-601, Coronado, Calif., November, 1976.

Duda R. O. and P. E. Hart, "Use of the Hough Transformation to Detect Lines and Curves in Pictures," Communications of the ACM, Vol. 15, No. 1, 11-15, January, 1972.

Duda R. O. and P. E. Hart, Pattern Classification and Scene Analysis, John Wiley and Sons, New York, 1973.

Dudani S. A., "An Experimental Study of Moment Methods for Automatic Recognition of Three-Dimensional Objects from Television Images," (Ph.D. Thesis), The Ohio State University, Columbus, Ohio, August, 1973.

Falk G., "Computer Interpretation of Imperfect Scenes Containing Polyhedra from Stereo Pairs of Views," Stanford Artificial Intelligence Laboratory Memo AIM-132, Stanford University, August, 1970.

Fletcher R., personal communication.

Fuchs H., Duran J., and B. Johnson, "A System for Automatic Acquisition of Three-Dimensional Data," Proceedings of the 1977 National Computer Conference, A. F. I. P. S., 49-53, Dallas, Texas, June 13-16, 1977.

Ganapathy S., "Reconstruction of Scenes Containing Polyhedra from Stereo Pairs of Views," Stanford Artificial Intelligence Laboratory Memo AIM-272, Stanford University, December, 1975.

Graham R. E., "Snow Removal: A Noise Stripping Process for Picture Signals," I. R. E. Trans. on Information Theory, Vol. IT-8, 129-144, 1962.

Grape G. R., "Model Based (Intermediate-Level) Computer Vision," Stanford Artificial Intelligence Laboratory Memo AIM-201, Stanford University, May, 1973.

Guzman A., "Computer Recognition of Three-Dimensional Objects in a Visual Scene," (Thesis) MAC-TR-59, M. I. T., December, 1968.

Hannah M. J., "Computer Matching of Areas in Stereo Images," Stanford Artificial Intelligence Laboratory Memo

AIM-239, Stanford University, July, 1974.

Hemami H., Weimer F. C. and J. G. Advani, "Identification of Three-Dimensional Objects by Sequential Image Matching," Proceedings of the Conference on Computer Graphics, Pattern Recognition and Data Structures, 273-278, UCLA, May 1975.

Horn B. K. P., "Shape from Shading: A Method for the Determination of the Shape of a Smooth Opaque Object from One View," (Thesis), Artificial Intelligence Report TR-79, Electrical Engineering Dept., M. I. T., Cambridge, Mass., 1970.

Horowitz S. L. and T. Pavlidis, "Picture Segmentation by a Directed Split and Merge Procedure," Proceedings of the Second International Joint Conference on Pattern Recognition, 424-433, Copenhagen, 1974.

Idesawa M., Yatagai T. and T. Soma, "A Method for Automatic Measurement of Three-Dimensional Shape by a New Type of Moire Topography," Proceedings of the Third International Joint Conference on Pattern Recognition, 708-712, Coronado, Calif., November, 1976.

Jarvis J. F., "Feature Recognition in Line Drawings Using Regular Expressions," Proceedings of the Third International Joint Conference on Pattern Recognition, 189-192, Coronado, Calif., November, 1976.

Jones V. C., "Tracking: An Approach to Dynamic Vision and Hand-Eye Coordination," Coordinated Science Laboratory Report R-696, U. of Illinois, Urbana, December, 1975.

Lev A., Zucker S. W. and A. Rosenfeld, "Iterative Enhancement of Noisy Images," I. E. E. E. Trans. on Systems, Man, and Cybernetics, Vol. SMC-7, No. 6, 435-442, June 1977.

Levine M. D., O'Handley D. A. and G. M. Yagi, "Computer Determination of Depth Maps," Computer Graphics and Image Processing, Vol. 2, 131-150, 1973.

Lowerre B. T., "The HARPY Speech Recognition System," (Ph.D. Thesis), Computer Science Dept., Carnegie-Mellon University, Pittsburgh, Pennsylvania, 1976.

Markarian H., Bernstein R., Ferneyhough D. G., Gregg L. E. and F. S. Sharp, "Digital Correction for High-Resolution Images," Photogrammetric Engineering, 1311-1320, 1973.

Marr D. and H. K. Nishihara, "Representation and Recognition of the Spatial Organization of Three-Dimensional Shapes," M. I. T. Artificial Intelligence Memo 377, August, 1976.

Marr D. and T. Poggio, "Cooperative Computation of Stereo Disparity," Science, Vol. 194, 283-287, 1976.

Mckee J. W. and J. K. Aggarwal, "Computer Recognition of Partial Views of Three-Dimensional Curved Objects," Information Systems Research Laboratory Technical Report No. 171, University of Texas at Austin, May, 1975.

Montanari U., "A note on Minimal Length Polygonal Approximation to a Digitized Contour," Communications of the ACM, Vol. 13, 41-47, January, 1970.

Mori K., Kidode M. and H. Asada, "An Iterative Prediction and Correction Method for Automatic Stereo Comparison," Computer Graphics and Image Processing, Vol. 2, 393-401, 1973.

Nahi N. E. and A. Habibi, "Decision Directed Recursive Image Enhancement," I. E. E. E. Trans. on Circuits and Systems, Vol. CAS-22, No. 3, 286-293, March, 1975.

Nevatia R., "Structured Descriptions of Complex Curved Objects for Recognition and Visual Memory," Stanford Artificial Intelligence Laboratory Memo AIM-250, October, 1974.

Nevatia R., "Depth Measurement by Motion Stereo," Computer Graphics and Image Processing, Vol. 5, 203-214, 1976.

O'Handley D. A., "Scene Analysis in Support of a Mars Rover," Computer Graphics and Image Processing, Vol. 2, 281-297, 1973.

Pavlidis T. and F. Ali, "Contour Description by General Syntactic Techniques," Proceedings of the Workshop on Picture Data Description and Management, 16-2?, Chicago, Illinois, April, 1977.

Pavlidis T. and S. L. Horowitz, "Segmentation of Plane Curves," I. E. E. E. Trans. on Computers, Vol. C-23, No. 8, August, 1974.

Perkins D. N., "Computer Stereo Vision, A Combinatorial Theory with Implementation," (Ph.D. Thesis), Department of Mathematics, M. I. T., Cambridge, Mass., June, 1970.

Perkins W. A., "Model-Based Vision System for Scenes Containing Multiple Parts," Proceedings of the Fifth International Joint Conference on Artificial Intelligence, 678-684, M. I. T., Cambridge, Mass., 1977.

Pingle K. K. and A. J. Thomas, "A Fast, Feature-Driven Stereo Depth Program," Stanford Artificial Intelligence Laboratory Memo AIM-248, Stanford University, May, 1975.

Quam L. H., "Computer Comparison of Pictures," Stanford Artificial Intelligence Laboratory Memo AIM-144, Stanford University, 1971.

Quam L. H. and M. J. Hannah, "Stanford Automatic Photogrammetry Research," Stanford Artificial Intelligence Laboratory Memo AIM-254, Stanford University, December, 1974.

Rabinowitz A. D., "Reconstruction of Polyhedra from Sets of Their Perspective Projections," (Ph. D. Thesis), Department of Electrical Engineering, New York University, April, 1971.

Ramer U., "An Iterative Procedure for the Polygonal Approximation of Plane Curves," Computer Graphics and Image Processing, Vol. 1, No. 3, 244-256, November, 1972.

Roberts L. G., "Machine Perception of Three-Dimensional Solids," Optical and Electro-Optical Information Processing, (ed. Tippett et al.), 159-197, M. I. T. Press, Cambridge, Mass., 1965.

Rocker F., "Localization and Classification of Three-Dimensional Objects," Second International Joint Conference on Pattern Recognition, 527-528, Copenhagen, August, 1974.

Rosenfeld A., Hummel R. A. and S. Zucker, "Scene Labeling by Relaxation Operations," I. E. E. E. Trans. on Systems, Man, and Cybernetics, Vol. SMC-6, No. 6, 420-433, June, 1976.

Rosenfeld A. and L. Davis, "Hierarchical Relaxation," Advance Papers of the Workshop on Computer Vision, University of Massachusetts, Amherst, June, 1977.

Rubin S. M. and R. Reddy, "The Locus Model of Search and Its Use in Image Interpretation," Proceedings of the Fifth International Joint Conference on Artificial Intelligence, 590-595, M. I. T., Cambridge, Mass., August, 1977.

Schnier M. O., "Recognition Using Semantic Constraints," Proceedings of the Fifth International Joint Conference on Artificial Intelligence, 585-589, M. I. T., Cambridge, Mass., August, 1977.

Shapira R. "A Technique for the Reconstruction of a Straight-Edge, Wire-Frame Object from Two or More Plane Projections," Computer Graphics and Image Processing, Vol. 3, 318-326, 1974.

Shapira R. and H. Freeman, "Reconstruction of Curved Surface Bodies from a Set of Imperfect Projections," Proceedings of the Fifth International Joint Conference on

Artificial Intelligence, 628-634, M. I. T., Cambridge, Mass., August, 1977.

Shirai Y., "Edge Finding, Segmentation of Edges and Recognition of Complex Objects," Proceedings of the Fourth International Joint Conference on Artificial Intelligence, 674-681, Tbilisi, Georgia, U. S. S. R., August, 1975.

Sklansky J., Chazin R. L. and B. J. Hansen, "Minumum-Perimeter Polygons of Digitized Silhouettes," I. E. E. E. Trans. on Computers, Vol. C-21, No. 3, 260-268, March, 1972.

Tasto M. and U. Block, "Locating Objects in Complex Scenes Using a Spatial Distance Measure," Proceedings of the Second International Joint Conference on Pattern Recognition, 336-340, Copenhagen, August, 1974.

Tsuji S. and F. Matsumoto, "Detection of Elliptic and Linear Edges by Searching Two Parameter Spaces," Proceedings of the Fifth International Joint Conference on Artificial Intelligence, 700-705, M. I. T., Cambridge, Mass., August, 1977.

Turner K. J., "Computer Perception of Curved Objects Using a Television Camera," (Thesis), U. of Edinburgh, 1974.

Underwood S. A. and C. L. Coates, "Visual Learning from Multiple Views," I. E. E. E. Trans. on Computers, Vol. C-24, No. 6, 651-661, 1975.

Vanderbrug G. J. and A. Rosenfeld, "Two Stage Template Matching," I. E. E. E. Trans. on Computers, Vol. C-26, No. 4, 384-393, April, 1977.

Waltz D. L., "Generating Semantic Descriptions from Drawings of Scenes with Shadows," M. I. T. Artificial Intelligence Memo AI-TR-271, M. I. T., Cambridge, Mass., November, 1972.

Wong K. W., Gamble E. V. and R. E. Riggins, "Analysis of RBV Television System," Photogrammetric Engineering, 851-861, 1973.

Zucker S. W., Hummel R. A. and A. Rosenfeld, "An Application of Relaxation Labeling to Line and Curve Enhancement," I. E. E. E. Trans. on Computers, Vol. C-26, No. 4, 394-403, April, 1977.

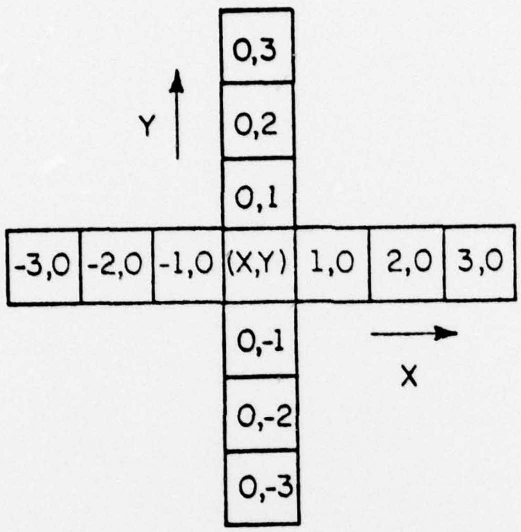
Zahn C. T., "Graph-Theoretical Methods for Detecting and Describing Gestalt Clusters," I. E. E. E. Trans. on Computers, Vol. C-20, No. 1, 68-86, January, 1971.

APPENDIX

A. Edge Detection and Tracking.

The operator used is a simple difference over a cross pattern (Figure A-1). Whenever the difference exceeds a threshold, a test is performed to determine whether or not the gradient is at a local maximum with respect to the x- or y-axis directions. If so, then it is retained as a valid edge point and its eight nearest neighbors are searched to find an additional edge satisfying the same criteria. In this way edges are immediately tracked without intermediate storage and thinning of an edge picture. When a significant nearest neighbor is not found, then next-nearest neighbors are searched and so on. If third neighbors show no edge, then the tracking is terminated. Gradient position, angle, and intensity are stored on an output list in the sequence in which they are found.

There is some prejudice in tracking an edge in the direction in which the last edge was found due to the nature of the algorithm. The nearest neighbors to the next to last found edge are effectively erased, so that further search is restricted to a fan beam in the direction of the edge contour (see Figure A-2). This is desirable since it prevents detection of sharp bends of noise in the curve. Further narrowing of the fan beam might not be desirable at this level, since there exists noise in the picture, and a



$$DX = \sum_{i=1}^{3} \left[I(X+i,Y) - I(X-i,Y) \right]$$

$$DY = \sum_{i=1}^{3} \left[I(X,Y+i) - I(X,Y-i) \right]$$

$$\text{Magnitude} = \sqrt{DX^2 + DY^2}$$

$$\text{Angle} = \operatorname{Tan}^{-1}\,(DX/DY)$$

FP-5736

Figure A-1 Cross mask used in edge detection. Three pixels are averaged on each side of center to provide some noise cancellation and resolution enhancement.

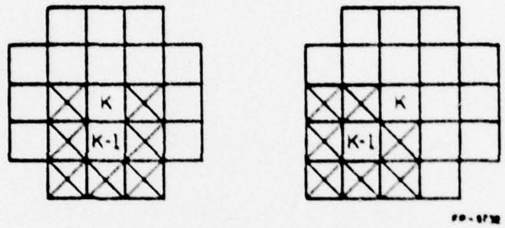


Figure A-2 Illustration of local erasure to prevent re-detection of an edge chain. K and K-1 correspond to the current and last visited edge locations. Before searching about K for the next edge element, a neighborhood about K-1 is effectively erased (X's). This has an added effect of restricting search for the next point to a fan beam in the general direction of contour growth. Blank pixels correspond to these search locations.

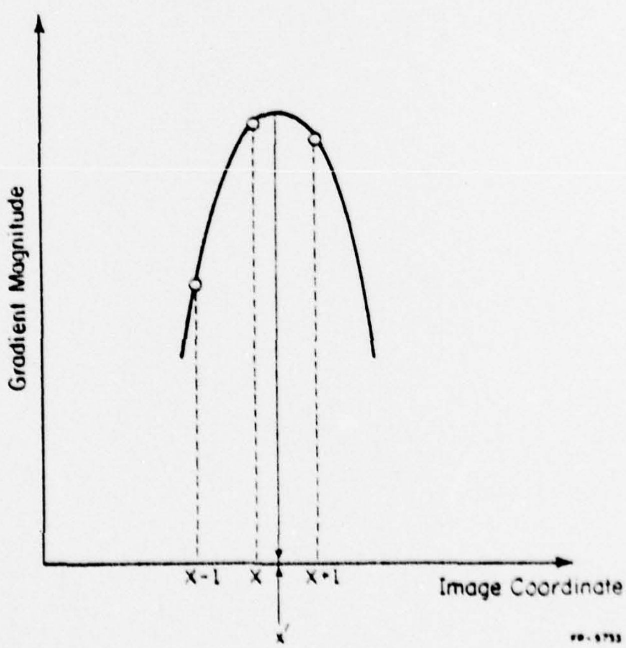


Figure A-3 Illustration of inter-pixel parabolic interpolation for edge position refinement. Grey level resolution can essentially be translated into positional enhancement provided that one can make assumptions about the nature of the edge shape at its center.

broad fan is often required to maintain continuity of edge following. Restricting the gradient algorithm to local peaks constrains the edge movement sufficiently so that further narrowing of the fan beam is not needed.

Edge position is improved in precision by parabolic interpolation as illustrated in Figure A-3. The refined position, x' , is defined by

$$x' = x + INC, \tag{9}$$

where

$$INC = \frac{(g3 - g1)}{4(g2 - g1/2 - g3/2)}, \tag{10}$$

and g1, g2, and g3 are the gradient values at three successive coordinates in the picture (x or y directions). When g2 is a local extremum (a gradient peak), INC takes on values between -0.5 and +0.5.

VITA

David Joseph Burr was born in Canonsburg, Pennsylvania, on June 15, 1946. He received the B.A. degree with honors in physics from Franklin and Marshall College in 1968, with a thesis entitled "The Mossbauer Effect in an Iron-Germanium Alloy". He entered graduate school in physics at the University of Illinois in 1968, as a Teaching Assistant. He entered service in the United States Air Force in 1969 and worked as a Scientific Assistant at Cape Kennedy. During 1971 he completed M.S. degree requirements at the Antenna Laboratory, University of Illinois. Upon discharge from the service in 1973 he returned to the University of Illinois, Coordinated Science Laboratory, as a University Fellow. There he received the Ph.D. degree for his work in computer vision. Additional experience includes two summers in industrial research laboratories.

While attending the University of Illinois he coauthored four papers entitled "Remote Sensing of Complex Permittivity by Multipole Resonances in RCS," "The Minimal Spanning Tree in Visual Data Segmentation," "A System for Stereo Computer Vision with Geometric Models," and "Multiple View Stereo Vision with Wire Frame Models" (submitted for publication). He is a member of Phi Beta Kappa, Phi Kappa Phi, Sigma Pi Sigma, Pi Mu Epsilon, I. E. E. E., and the Acoustical and Audio Engineering Societies.